

c	32	33.4	1.7	5001	6	US-10-517-441-175	Sequence 417, App
	33	33.4	1.7	5493	6	US-10-517-441-175	Sequence 417, App
	34	33.4	1.7	8900	6	US-10-517-441-428	Sequence 428, App
	35	33.4	1.7	8900	6	US-10-517-441-702	Sequence 702, App
	36	33.4	1.7	16579	6	US-10-517-441-564	Sequence 564, App
	37	33.4	1.7	19634	6	US-10-517-441-413	Sequence 413, App
	38	33.4	1.7	1552	6	US-11-217-529-166165	Sequence 166165, App
	39	35.2	1.7	1056	6	US-11-217-529-166165	Sequence 166165, App
	40	35.2	1.7	4022	6	US-10-517-441-732	Sequence 732, App
	41	58.8	1.7	8900	6	US-10-517-441-701	Sequence 701, App
	42	58.8	1.7	5026	6	US-11-217-529-166180	Sequence 166180, App
	43	58.8	1.7	6731	6	US-10-517-441-747	Sequence 747, App
	44	58.8	1.7	8666	6	US-10-517-441-284	Sequence 284, App
	45	58.6	1.7	4022	6	US-10-517-441-458	Sequence 458, App
	46	58.6	1.7	12423	6	US-10-517-441-526	Sequence 526, App

ALIGNMENTS

APPLICANT: ASHIKARI, TOSHIHIKO  
TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
FILE REFERENCE: 5-38-285  
CURRENT APPLICATION NUMBER: US/11/217,529  
CURRENT FILING DATE: 2005-09-02  
PRIOR APPLICATION NUMBER: US/10/932,182  
PRIORITY DATE: 2004-09-02  
NUMBER OF SEQ ID NOS: 197023  
SOFTWARE: PatentIn version 3.3

```

; SEQ ID NO 3631
; LENGTH: 2571
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus

```

7000-570-17-11-50

Query match	Similarity	60.0%; Best local	Score: 8.18e-65;	Indels	162;	Gaps	2
Matches	823;	Conservative	0;	Mismatches	387;		
1606	QOY	AGCAGAAATTTCTGATGGGATTAATGAGAGTTTCAATACCTGATAGATATTTTCACATTTTTC	1666				
1185	DBb	AGCATATTTGTCTACGAGAAATACAGATATTTCAATATCCCAACAGATTTTCAATTTTTC	1242				
1666	QOY	TTCTGAATCAGAGAAACTATTCATCTGCCCCGATATTTCCATCATATAGTATCACCAAGTCA	1721				
1245	DBb	TTCAGAACTCGATCAAAACCCGCTCATGCTAGTGATATTTCCGTCCTCGGTGTGTGAGGGTCA	1301				
1726	QOY	ATCTGTTCGAGATTTATTTTGAATATGCTGAGAAACTTGTGTGGTTAGATTTGTTACTTTGTCC	1787				
1305	DBb	GTCATTTCTATGAAATTTGTTCTCAGAGGGGGGAACCAACTTGTGTGGTTAGATTGCTAGTTGCC	1361				
1786	QOY	TACTGATTCGGAAATGAAATGTTTGGGCCAAAGCAATTTGGTGTATTCATCTCTTTAACTGCTCA	1848				
1365	DBb	AACCGATGACGAAATGCGTTGCTATAGCAAAAGCTTTTGGAAATTCATTCATTGACGGCAGA	1421				
1846	QOY	AGATATTTTCGAATCTCAAGAACTCGTGAAGAAAGCTGTGAATATTTATTAAGTTATTTATTTGT	1901				
1425	DBb	AGATATTTAGAAATGCAAGAAATCTCTGTGAAGAAATGAGCTTTTTTAAGCTGTATTTATTTTGT	1488				
1906	QOY	TTGTTTCCATCTTTTGAAGCTGATTAAGAAATCTGAAGATATTTTAGAACCGATAAATTTT	1969				
1485	DBb	TTGTTTCCATACATTTTGAAGAACGCAAGAAATCTGAAGATATTTCTTAGAGCCCAATTAACGT	1546				
1966	QOY	TTATATTTGTTTCTTTCATGATGATATTAACAGTTCCTCAATTTTTCACCAATTTTCTCATCC	2028				
1545	DBb	TTACATTTGCTGTTTGTAGATCTGGTGTTTTGAACCTTCCATTTTGTGTCCAAATTTTCACAT	1608				
2026	QOY	AGCAAAATGTTTAGAAGAAGATTTCTGCTCAATTTTAGAGAGATTAATGTCTGATGTTAGTCTCAT	2088				
1605	DBb	TGCTAATATGTTAGAAGACGTGTGAAGCAAACTACGACAGATTAACGTCAACGCTCAATTTCCGA	1668				

OM nucleic - nucleic search, using sw model

Run on: June 15, 2006, 03:46:50 ; Search time 3642 Seconds  
(without alignments)  
11892.892 Million cell updates/sec

Title: US-10-018-105A-3

Perfect score: 323

Sequence: 1 tatataataataatata.....tcaattattatcaattgc 3525

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database :
- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
  - 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
  - 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
  - 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
  - 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
  - 16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2769	78.6	2769	7	US-10-032-585-6675
2	408.8	11.6	2571	11	US-10-932-182A-5831
3	403.8	11.5	2574	11	US-10-932-182A-2281
4	396	11.2	2580	11	US-10-932-182A-78462
5	383	10.9	2577	11	US-10-932-182A-76424
6	179.6	5.1	1992	6	US-10-128-714-2563
7	179.6	5.1	1992	6	US-10-128-714-2563
8	127.4	3.6	2910	11	US-10-932-182A-77545
9	118	3.3	462	3	US-09-864-408A-2265
10	117.6	3.3	8056	9	US-10-473-126-386
11	116.4	3.3	688	9	US-10-653-047-7103
12	116.2	3.3	2910	11	US-10-932-182A-3077
13	113.6	3.2	2048	6	US-10-128-714-1563
14	113.6	3.2	2048	6	US-10-128-714-1563
15	113.6	3.2	4048	6	US-10-128-714-563
16	113.6	3.2	4048	6	US-10-128-714-563
17	106.8	3.0	3089	10	US-10-954-778-160
18	106.8	3.0	4985	6	US-10-094-240-10
19	106.8	3.0	4985	6	US-10-056-405-10
20	106.2	3.0	575	9	US-10-653-047-447
21	97.4	2.8	286	9	US-10-425-115-137269
22	94.4	2.7	969	9	US-10-363-483A-26081
23	94.4	2.7	969	9	US-10-363-483A-26082

ALIGNMENTS

RESULT 1

US-10-032-585-6675  
; Sequence 6675, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6675  
; LENGTH: 2769  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-10-032-585-6675

Query Match 78.6%; Score 2769; DB 7; Length 2769;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2769; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	338	ATGTCGATAGTGAAGTATATCAAAATTCACCTATCACTATCACTATTCCTAGATCT	397
DB	1	ATGTCGATAGTGAAGTATATCAAAATTCACCTATCACTATTCCTAGATCT	60
QY	398	GATGAAGTATGATATCAATCAAAATTCACCTATTCCTAGATCT	457
DB	61	GATGAAGTATGATATCAATCAAAATTCACCTATTCCTAGATCT	120
QY	458	GATGAAGTATGATATCAATCAAAATTCACCTATTCCTAGATCT	517
DB	121	GATGAAGTATGATATCAATCAAAATTCACCTATTCCTAGATCT	180
QY	518	CAACATCATCATGATATCAATCAAAATTCACCTATTCCTAGATCT	577
DB	181	CAACATCATCATGATATCAATCAAAATTCACCTATTCCTAGATCT	240
QY	578	TCTTCAATTAAGAAAAATTCATCTTACCGATTAAGATTAACCACTATTCCTAGATCT	637
DB	241	TCTTCAATTAAGAAAAATTCATCTTACCGATTAAGATTAACCACTATTCCTAGATCT	300
QY	638	TTATTCGTGATGATATCAATCAAAATTCACCTATTCCTAGATCT	697
DB	301	TTATTCGTGATGATATCAATCAAAATTCACCTATTCCTAGATCT	360
QY	698	TCAATACGTAAGATATTTTATTTTAAAGATTAATCTGACGACAAATTTCTACTAATATCAT	757
DB	361	TCAATACGTAAGATATTTTATTTTAAAGATTAATCTGACGACAAATTTCTACTAATATCAT	420



GenCore version 5.1.9  
 Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 15, 2006, 03:11:16 ; Search time 653 Seconds  
 (without alignments)  
 10100.549 Million cell updates/sec

Title: US-10-018-105A-3  
 Perfect score: 3523  
 Sequence: 1 taktataataataataata.....tcaattattatcaatgtgc 3525

Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database :

Issued\_Patents\_NA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1\_COMB.seq:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5\_COMB.seq:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A\_COMB.seq:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B\_COMB.seq:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7\_COMB.seq:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/8\_COMB.seq:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/9\_COMB.seq:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/10\_COMB.seq:\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/11\_COMB.seq:\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/12\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2155.2	61.2	2358	3	US-09-248-796A-6328
2	423.8	12.0	507	3	US-09-248-796A-6327
3	188.2	4.8	537	3	US-08-998-416-66
4	118.2	3.4	2445	3	US-09-248-796A-6302
5	116.4	3.3	688	3	US-09-533-559-7103
6	106.2	3.0	575	3	US-09-533-559-447
7	74.4	2.1	50000	3	US-09-662-254B-23
8	73.4	2.1	708	3	US-09-248-796A-10622
9	71.8	2.0	19124	2	US-08-487-826B-13
10	70.4	2.0	50000	3	US-09-662-254B-25
11	70.2	2.0	18651	3	US-09-949-002-592
12	70.2	2.0	18682	3	US-09-949-002-786
13	68.8	2.0	1141	3	US-09-806-708B-22
14	68.2	1.9	50000	3	US-09-662-254B-24
15	68	1.9	767677	3	US-09-949-016-12147
16	67.6	1.9	767677	3	US-09-949-016-17361
17	67.6	1.9	29717	3	US-09-949-016-16284
18	67.6	1.9	32392	3	US-09-662-254B-27
19	67.6	1.9	60376	3	US-09-949-016-12423
20	66.8	1.9	1410	3	US-09-662-254B-33
21	66.8	1.9	14066	3	US-09-601-198-56
22	66.4	1.9	19438	3	US-09-949-016-12699
23	66	1.9	26000	3	US-09-843-376-10
24	65.8	1.9	205044	3	US-09-949-016-15851
25	65.8	1.9	205044	3	US-09-949-016-15852
26	65.8	1.9	205044	3	US-09-949-016-15853
27	65.8	1.9	223471	3	US-09-949-016-12387
28	65.8	1.9	223471	3	US-09-949-016-12724
29	65.8	1.9	223471	3	US-09-949-016-12725

30	65.6	1.9	50000	3	US-09-662-254B-23	Sequence 23, App1
31	65.4	1.9	32392	3	US-09-662-254B-27	Sequence 27, App1
32	65.2	1.9	1037	3	US-09-181-585-3	Sequence 3, App1
33	65.2	1.9	1159	3	US-09-181-585-1	Sequence 1, App1
34	65.2	1.9	1471	3	US-09-181-585-2	Sequence 2, App1
35	65	1.8	700	3	US-09-735-271-1038	Sequence 1038, App
36	65	1.8	18773	3	US-09-949-016-14164	Sequence 14164, A
37	64	1.8	1141	3	US-09-806-708B-22	Sequence 22, App1
38	63.6	1.8	50000	3	US-09-662-254B-26	Sequence 26, App1
39	63.6	1.8	119153	3	US-09-949-016-12378	Sequence 12378, A
40	63.4	1.8	837	3	US-08-998-416-288	Sequence 288, App
41	63.4	1.8	187169	3	US-09-949-016-12776	Sequence 12776, A
42	63.4	1.8	191569	3	US-09-949-016-15940	Sequence 15940, A
43	63.2	1.8	50000	3	US-09-662-254B-25	Sequence 25, App1
44	62.4	1.8	114139	3	US-09-949-016-16536	Sequence 16536, A
45	62	1.8	4441	3	US-09-796-348-20	Sequence 20, App1

## ALIGNMENTS

RESULT 1  
 US-09-248-796A-6328  
 ; Sequence 6328, Application US/09248796A  
 ; Patent No. 6747137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Keith Weinstock et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248.796A  
 ; PRIOR FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: US 60/074.725  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096.409  
 ; PRIOR FILING DATE: 1998-08-13  
 ; NUMBER OF SEQ ID NOS: 28208  
 ; SEQ ID NO 6328  
 ; LENGTH: 2358  
 ; TYPE: DNA  
 ; ORGANISM: Candida albicans  
 US-09-248-796A-6328

Query Match 61.2%; Score 2155.2; DB 3; Length 2358;  
 Best Local Similarity 97.2%; Pred. No. 0;  
 Matches 2293; Conservative 0; Mismatches 23; Indels 44; Gaps 8;

QY	285	ATAAATAAATTAAGAGTATATCTCCCTTTTGTGTTTTTCTTCCAGCCATGTCG	344
DB	1	ATAAATAAATTAAGAGTATATCTCCCTTTTGTGTTTTTCTTCCAGCCATGTCG	58
QY	345	ATAGTGAAGTTTATTAATCAAAATCACTACTAATCACTATCTAGATCTGATGAAG	404
DB	59	ATAGTGAAGTTTATTAATCAAAATCACTACTAATCACTATCTAGATCTGATGAAG	118
QY	405	TATTGGATGATCATAGAAATCAAAATCACTAATGATTTGCCATAGTAGTGAAGATG	464
DB	119	TATTGGATGATCATAGAAATCAAAATCACTAATGATTTGCCATAGTAGTGAAGATG	178
QY	465	AGTTGGAAATTAATAATCAAGATTTAGAAATCAAGAGTTGTAAAGAGCAAAA	512
DB	179	AGTTGGAAATTAATAATCAAGATTTAGAAATCAAGAGTTGTAAAGAGCAAAA	238
QY	513	AACAAACACATCATCAAGAGATTCATCATAGATTAATGTAACCACTGCTGTAAT	572
DB	239	AACAAACACATCATCAAGAGATTCATCATAGATTAATGTAACCACTGCTGTAAT	298
QY	573	CTGGTTCTTCAATTAAGAAAAATCTAATCTTACCAGTAAGATAGTATACCAACCTA	632
DB	299	CTGGTTCTTCAATTAAGAAAAATCTAATCTTACCAGTAAGATAGTATACCAACCTA	358
QY	633	TCAGTTATCTGGTGGTGTATGTAATAATTAACACCGGTCAAAAAATCGTAAATTAACA	692
DB	359	TCAGTTATCTGGTGGTGTATGTAATAATTAACACCGGTCAAAAAATCGTAAATTAACA	418
QY	693	TCAGTTATCTGGTGGTGTATGTAATAATTAACACCGGTCAAAAAATCGTAAATTAACA	752
DB	419	TCAGTTATCTGGTGGTGTATGTAATAATTAACACCGGTCAAAAAATCGTAAATTAACA	478

Db		1619	AAAAAGTGAATATTATTTAAAGGTATTATTTGCTGGTTCCTCACTCTTTTGAAAGCTGATA	1678
QY		1932	AAGAATCTCAGAGATTAATTTAGAACCGATAAATCTTTATATATGTGTGTTCCATCATATGCTA	1991
Db		1679	AGAACTCGAGATATTTAGAACCGATAAATCTTTATATATGTGTGTTCCATCATATGCTA	1738
QY		1992	TATTTACAGTCCCAATTTTCCAGATTTCTCCATCCACGAAATGTTATGAAGAAGATGTCCTC	2051
Db		1739	TATTTACAGTCCCAATTTTCCAGATTTCTCCATCCACGAAATGTTATGAAGAAGATGTCCTC	1798
QY		2052	AATTTAGAGAGATTATGCCAGTTTAGTGCCTGATTTGCTATATCTATATGCTTAATCCGATGAAA	2111
Db		1799	AATTTAGAGAGATTATGCCAGTTTAGTGCCTGATTTGCTATATCTATATGCTTAATCCGATGAAA	1858
QY		2112	TTACCGATGGTTTTTSCCCCCGCTGATTTCAATGSAATGAATATGAAGCTGATGCCATTTGAAG	2171
Db		1859	TTACCGATGGTTTTTSCCCCCGCTGATTTCAATGSAATGAATATGAAGCTGATGCCATTTGAAG	1918
QY		2172	ATSCGCTTTTTCACCTGTATAGAGATCACTGATTTTAGTAGTAGTATGTTTACAAGAAATTTGCTGAAT	2231
Db		1919	ATSCGCTTTTTCACCTGTATAGAGATCACTGATTTTAGTAGTAGTATGTTTACAAGAAATTTGCTGAAT	1978
QY		2232	CAAGAAGAAAAAGTCAATGACTTTAAATGAGATTTATATCAGCTTAAAGCTGATGTCAATTAAAA	2291
Db		1979	CAAGAAGAAAAAGTCAATGACTTTAAATGAGATTTATATCAGCTTAAAGCTGATGTCAATTAAAA	2038
QY		2292	TGTTTGTCTAAAGATGTCGAAGAAGCTTAATTCCTTCTCTGCTTATTTATCAACGCTCAAT	2351
Db		2039	TGTTTGTCTAAAGATGTCGAAGAAGCTTAATTCCTTCTCTGCTTATTTATCAACGCTCAAT	2098
QY		2352	ATTAATCTTACCAACAACAACAACAACA-----GGCCCCACCACCAACCTCAATCTCTTATTA	2405
Db		2099	ATTAATCTTACCAACAACAACAACAACAACAAGCCCCACCACCACCTTAACCTTATTA	2158
QY		2406	TTACTTCCAACAATTAATTTCAAACCTTTGAATCTTAAATAGTTTATGAAATCTTCAACTGCTGGAG	2465
Db		2159	TTACTTCCAACAATTAATTTCAAACCTTTGAATCTTAAATAGTTTATGAAATCTTCAACTGCTGGAG	2218
QY		2466	GAGTAGGAGTAGAGGAAATTAATTTTGGTCCCACCTCCCACTSGAAAAATAATPACTAATACTA	2525
Db		2219	GAGTAGGAGTAGAGGAAATTAATTTTGGTCCCACCTCCCACTSGAAAAATAATPACTAATACTA	2278
QY		2526	ATATCTAATPACTACTGGTTCACCTTCAACGCTCAACGCTCAACCAACAACAACATPGTATCACTA	2585
Db		2279	ATATCTAATPACTACTGGTTCACCTTCAACGCTCAACGCTCAACCAACAACAACATPGTATCACTA	2335
QY		2586	ACAAATCTTCCGATCCCC	2605
Db		2336	ACAAATCTTCCGATCCCC	2355
 RESULT 2 US-09-248-796A-6327 ; Sequence 6327 Application US/09248796A ; Patent No. 6747137 ; GENERAL INFORMATION: ; APPLICANT: Keith Weinstein et al ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS				
			; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS	
			; PILE REFERENCE: 107196.132	
			; CURRENT APPLICATION NUMBER: US/09/248,796A	
			; CURRENT FILING DATE: 1999-02-12	
			; PRIOR FILING DATE: US 60/074,725	
			; PRIOR FILING DATE: 1998-02-13	
			; PRIOR APPLICATION NUMBER: US 60/096,409	
			; PRIOR FILING DATE: 1998-08-13	
			; NUMBER OF SEQ ID NOS: 28208	
			; SEQ ID NO 6327	
			; LENGTH: 507	
			; TYPE: DNA	
			; ORGANISM: Candida albicans	
			US-09-248-796A-6327	
 Query Match                    12.0%; Score 423.8; DB 3; Length 507; Best Local Similarity       98.0%; Pred. No. 4.2e-73; Matches 440; Conservative     0; Mismatches     7; Indels       2; Gaps       1;				
QY		2660	ATAATCAACCAATGTTTCAAAATTTATAG--CCTATGAAAAAATTTTCAGCTGTTTCACATT	2717

OM protein - protein search, using sw model

Run on: June 13, 2006, 14:21:02 ; Search time 17 Seconds  
(without alignments)  
685.708 Million cell updates/sec

Title: US-10-018-105A-4  
Perfect score: 4820  
Sequence: 1 MSDESQYQNTNQIPRS.....LGLKHGNGKSIIFPNKYE 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 64916 seqs, 12643201 residues

Total number of hits satisfying chosen parameters: 64916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New:

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US09\_NEW\_PUB pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US06\_NEW\_PUB pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US07\_NEW\_PUB pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US08\_NEW\_PUB pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/PCT\_NEW\_PUB pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US10\_NEW\_PUB pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US11\_NEW\_PUB pep.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match %	Length DB	ID	Description
1	166.5	3.5	271	6	US-10-471-571A-3948
2	157	3.3	750	7	US-11-134-228A-32
3	137	2.8	454	7	US-11-134-228A-35
4	137	2.8	1392	6	US-10-953-349-1430
5	136	2.8	315	6	US-10-471-571A-4384
6	136	2.8	2478	6	US-10-471-571A-2278
7	135.5	2.8	596	6	US-10-196-749-310
8	135.5	2.8	596	6	US-11-101-316-100
9	133	2.8	933	7	US-11-121-154-178
10	130	2.7	1274	6	US-10-471-571A-898
11	129.5	2.7	624	6	US-10-528-104-3
12	128.5	2.7	488	6	US-10-471-571A-10
13	128.5	2.7	788	7	US-11-293-697-4150
14	126	2.6	710	7	US-11-134-228A-33
15	124.5	2.6	477	7	US-11-134-228A-34
16	124.5	2.6	935	6	US-10-471-571A-4496
17	123	2.6	688	7	US-11-293-697-2798
18	121	2.5	940	7	US-11-293-697-3781
19	119.5	2.5	774	6	US-10-471-571A-4824
20	119.5	2.5	791	7	US-11-121-154-24
21	119.5	2.5	1106	7	US-11-293-697-4301
22	119	2.5	502	6	US-10-511-937-3008
23	118	2.4	387	6	US-10-953-349-8568
24	118	2.4	439	6	US-10-953-349-8567
25	117	2.4	1003	6	US-10-953-349-1431
26	116.5	2.4	2937	6	US-10-480-962-18
27	116.5	2.4	2969	6	US-10-480-962-19
28	116	2.4	1075	6	US-11-134-228A-23
29	115.5	2.4	1261	6	US-10-471-571A-1304
30	113	2.3	611	6	US-10-953-349-3922
31	113	2.3	624	6	US-10-953-349-3921

32	113	2.3	877	6	US-10-471-571A-3884	Sequence 3884, Ap
33	111.5	2.3	560	6	US-10-953-349-3923	Sequence 3923, Ap
34	111.5	2.3	1349	6	US-10-471-571A-3352	Sequence 3352, Ap
35	111	2.3	340	6	US-10-953-349-20922	Sequence 20922, A
36	111	2.3	613	7	US-11-293-697-4125	Sequence 4125, A
37	111	2.3	619	6	US-10-471-571A-3914	Sequence 3914, Ap
38	111	2.3	895	6	US-10-471-571A-922	Sequence 922, App
39	111	2.3	2829	6	US-10-525-621-1	Sequence 1, Appl
40	110.5	2.3	568	6	US-10-953-349-10556	Sequence 10556, A
41	110.5	2.3	718	6	US-10-953-349-5389	Sequence 5389, Ap
42	110.5	2.3	738	6	US-10-953-349-5388	Sequence 5388, Ap
43	110.5	2.3	763	6	US-10-953-349-5387	Sequence 5387, Ap
44	110.5	2.3	1073	7	US-11-293-697-2915	Sequence 2915, Ap
45	110.5	2.3	1809	6	US-10-370-959-67	Sequence 67, Appl

### ALIGNMENTS

RESULT 1  
US-10-471-571A-3948  
; Sequence 3948, Application US/10471571A  
; Publication No. US20060115490A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026927WO  
; CURRENT APPLICATION NUMBER: US/10/471.571A  
; PRIOR FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: GB-0107661.1  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 5642  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 3948  
; LENGTH: 2271  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
; NAME/KEY: MISC\_FEATURE  
; FEATURE:  
; LOCATION: (1)..(2271)  
; OTHER INFORMATION: hypothetical protein  
US-10-471-571A-3948

Query Match 3.5%; Score 166.5; DB 6; Length 2271;  
Best Local Similarity 18.3%; Pred. No. 0.052;  
Matches 109; Conservative 103; Mismatches 234; Indels 151; Gaps 20;

Qy	1	MSDESQYQNTNQIPRSDEVLDHNRQITNDCASIDSEDELEKSELESEVVKSEKQ	60
Db	1701	ISDSQSMSESVNDSVSESNSESDS--KMSGSTSVSDS--GSLSVSTSLRKSESVSESS	1757
Qy	61	QQRHQETTSNAPKPIVTRKS---GSSIKKSNLTKDRITNPKSLGGDDTIHSGHKNNRY	117
Db	1758	SLSCSQMSVSTVSDSSSVSTSLKSESVSESDSLSDSKSTSGSTSTSGSLSTST	1817
Qy	118	NMS---SLRKDFYLKDN---TDDNSTNNHHLAIPPIPIPTPIIT-----NANKSR	163
Db	1818	SLSGSESVSESTSLSDSISNMSDSTSTSDSLSGSISLSGSTSLSTSDSLSDSKSESSQ	1877
Qy	164	RKSQLENPLPKKTKTGRNNNNFENDLV--SPMTKTKTNDSEDIITWTANMKKIG--	220
Db	1878	SMSGSESTSVSDSQSSSTSVSDSQSSSTSVSDSQSSSTSVSDSQSSSTSVSDSQSS	1937
Qy	221	---IGATTLGVGTGTAT-----ATATAAGR-----	244
Db	1938	SMSGSVSTVSTSLSDSISGTSVSDSSTSTSTSLSDSMSGOSQSTSTSLSGSTST	1997
Qy	245	-----RPSRSSIDSEADS-----HARRSQETEDVCFPMVGDHVRNG	283
Db	1998	SMSGSESTSVSTVSTSLSDSISGTSVSDSSTSTSTSLSDSMSGOSQSTSTSLSGST	2057
Qy	284	IDPDIDEFIRREAREAYLQOKIAKNILRAIDFQNLKNNITGCAHRPHYHHNNKK	343
Db	2058	SDSDS-----TSTSTSDSTSGST-----STSTSESL	2083
Qy	344	NNGGGGGSMALKYTPKNILKLTLSRPFTHENSSSSSEIYELKTKQPPKYDDQLS	403
Db	2084	STSGSGSTVSDTSMSESNSSSVMSQDKSDSTSIDSESV---STSTSTSLSDSTST	2140

## OM protein - protein search, using sw model

Run on: June 13, 2006, 14:08:32 ; Search time 52 Seconds  
(without alignments)  
1551.985 Million cell updates/sec

Title: US-10-018-105A-4

Perfect score: 4820

Sequence: 1 MSDESYQNSTNQIPRPS.....LGKKHGNGKSIISPPNKYE 922

Scoring table: BLOSUM62

Gap 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_Ah.\*

- 1: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/5\_COMB.pep.\*
- 2: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/6\_COMB.pep.\*
- 3: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/7\_COMB.pep.\*
- 4: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/8\_COMB.pep.\*
- 5: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/9\_COMB.pep.\*
- 6: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep.\*
- 7: /EMC\_Celetra\_SIDS3/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3785	78.5	2	US-09-248-796A-20431	Sequence 20431, A
2	1880	34.3	859	US-09-538-092-1371	Sequence 717, App
3	1734	17.3	814	US-09-248-796A-20405	Sequence 20405, A
4	716	14.3	188	US-09-248-796A-20430	Sequence 20430, A
5	256	5.2	326	US-09-502-540-11640	Sequence 11640, A
6	172.5	3.6	1051	US-09-538-092-469	Sequence 469, App
7	166.5	3.5	2283	US-10-172-502-4	Sequence 4, App1
8	166	3.4	719	US-09-248-796A-17559	Sequence 17559, A
9	161	3.3	366	US-09-712-363-193	Sequence 193, App
10	161	3.3	471	US-08-311-731A-168	Sequence 168, App
11	159.5	3.3	1177	US-09-134-001C-5106	Sequence 5106, App
12	159.5	3.3	1253	US-09-958-617A-18	Sequence 18, App1
13	159.5	3.3	1584	US-09-457-040B-27	Sequence 27, App1
14	159.5	3.3	2150	US-10-135-322-17	Sequence 17, App1
15	158.5	3.3	785	US-09-487-558B-348	Sequence 348, App
16	158	3.3	651	US-08-431-080-24	Sequence 24, App1
17	158	3.3	651	US-08-938-53A-24	Sequence 24, App1
18	158	3.3	651	US-09-345-294-24	Sequence 24, App1
19	157	3.3	858	US-09-248-796A-19055	Sequence 19055, A
20	154	3.2	1259	US-09-949-016-10366	Sequence 10366, A
21	153.5	3.2	982	US-09-248-796A-20628	Sequence 20628, A
22	152	3.2	2870	US-09-479-467A-15	Sequence 15, App1
23	152	3.2	2870	US-09-655-160-15	Sequence 15, App1
24	152	3.2	3178	US-09-479-467A-4	Sequence 4, App1
25	152	3.2	3178	US-09-655-160-4	Sequence 4, App1
26	151	3.1	315	US-09-710-279-2100	Sequence 2100, App
27	151	3.1	324	US-09-134-001C-4080	Sequence 4080, App
28	150	3.1	10182	US-09-134-001C-3159	Sequence 3159, App
29	147	3.0	676	US-09-134-001C-4318	Sequence 4318, App
30	147	3.0	404	US-09-710-279-2964	Sequence 2964, App
31	146.5	3.0	5024	US-09-710-279-2964	Sequence 2964, App
32	146.5	3.0	1444	US-09-949-016-9652	Sequence 9652, App

33	146.5	3.0	1667	US-09-270-767-41425	Sequence 41425, A
34	146.5	3.0	2004	US-09-538-092-1371	Sequence 1371, App
35	146.5	3.0	2004	US-09-949-016-6756	Sequence 6756, App
36	146	3.0	292	US-09-248-796A-25373	Sequence 25373, A
37	145	3.0	873	US-09-248-796A-20365	Sequence 20365, A
38	145	3.0	1435	US-08-568-459A-4	Sequence 4, App1
39	145	3.0	1435	US-08-487-826B-4	Sequence 4, App1
40	145	3.0	1435	US-09-210-288-4	Sequence 4, App1
41	145	3.0	1435	US-10-153-273-4	Sequence 4, App1
42	144.5	3.0	1085	US-08-431-080-28	Sequence 28, App1
43	144.5	3.0	1085	US-08-938-53A-28	Sequence 28, App1
44	144.5	3.0	1085	US-09-345-294-28	Sequence 28, App1
45	144.5	3.0	3092	US-09-487-558B-172	Sequence 172, App

## ALIGNMENTS

RESULT 1  
US-09-248-796A-20431  
; Sequence 20431, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/704,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20431  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-20431

Query Match 78.5%; Score 3785; DB 2; Length 785;  
Best Local Similarity 96.2%; Pred. No. 3e-296;  
Matches 740; Conservative 2; Mismatches 13; Indels 14; Gaps 7;

QY	1	MSDESYQNSTNQIPRPSDEVLDHNRQITNDCAISDSDELEKSELESEVAKSEK	59
DB	18	MSDESYQNSTNQIPRPSDEVLDHNRQITNDCAISDSDELEKSELESEVAKSEK	77
QY	60	---QOQHHEITSDNAKPLTKSGSIIKKSNLTKDRITNPMSLSGGDDTINSCHKRN	116
DB	78	QOQOQHHEITSDNAKPLTKSGSIIKKSNLTKDRITNPMSLSGGDDTINSCHKRN	137
QY	117	YNNSLRKDFYKXNDTNDSTNNHTLAIPITPIPTIITNANKSRKSOLENLPLIK	176
DB	138	YNNSLRKDFYKXNDTNDSTNNHTLAIPITPIPTIITNANKSRKSOLENLPLIK	197
QY	177	KKTIGRNNNNPNDLVSPMTGKTNDSEDTN-TSTTANMKLGIGATTLGVGTGTAT	235
DB	198	KKTIGRNNNNPNDLVSPMTGKTNDSEDTN-TSTTANMKLGIGATTLGVGTGTAT	257
QY	236	ATATATAA---GRPSRSSIDSEADSHASRSSQETBEDVCPFMVGDHVRVNGIDFDEIDEP	292
DB	258	ATATATAGGGRPSRSSIDSEADSHASRSSQETBEDVCPFMVGDHVRVNGIDFDEIDEP	317
QY	293	IREEREAYLQKQAKNLIIRIDFQNTSKNNTTSGASRPHYHHNNKNNKNGDG-GG	351
DB	318	IREEREAYLQKQAKNLIIRIDFQNTSKNNTTSGASRPHYHHNNKNNKNGDGSG	377
QY	352	SSMAALKYTPNLIKTLKLSRPFETHSSSSSEIYELKTKQPPKYDDQLSLTSTSTST	411
DB	378	SSMAALKYTPNLIKTLKLSRPFETHSSSSSEIYELKTKQPPKYDDQLSLTSTSTST	437
QY	412	--SGSGSQVKGFGARISDINGSLDPFSLFPHSESETHAPDIPSLVSPQSVROLF	469
DB	438	SGSGSQVKGFGARISDINGSLDPFSLFPHSESETHAPDIPSLVSPQSVROLF	497
QY	470	RNGEETWMLDCTCPTDSEMDLAKAFGHPHTAEDIRMOETREKVELFKSVYVCFHTE	529

34 165 3.4 4498 6 US-11-097-143-2577 Sequence 2577, Ap  
35 164.5 3.4 2271 4 US-10-282-122A-43924 Sequence 43924, A  
36 161 3.3 366 3 US-09-712-363-193 Sequence 193, App  
37 161 3.3 2719 5 US-10-732-923-8668 Sequence 8668, Ap  
38 160 3.3 373 2 US-08-945-749-9 Sequence 9, Appli  
39 159.5 3.3 1169 4 US-10-282-122A-71179 Sequence 71179, A  
40 159.5 3.3 1177 4 US-10-724-972A-6013 Sequence 6013, A  
41 159.5 3.3 1253 4 US-10-363-798-2 Sequence 2, Appli  
42 159.5 3.3 1253 6 US-11-185-924-18 Sequence 18, Appli  
43 159.5 3.3 1584 5 US-10-732-923-13548 Sequence 13548, A  
44 159.5 3.3 2150 4 US-10-135-322-17 Sequence 17, Appli  
45 159 3.3 2399 5 US-10-732-923-15036 Sequence 15036, A

ALIGNMENTS

RESULT 1  
US-10-032-585-7675  
; Sequence 7675, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032.585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7675  
; LENGTH: 922  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-10-032-585-7675

Query Match 99.9%; Score 4814; DB 4; Length 922;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 921; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSDESYQNSTNQIPRSDVLDIRNQITNDCAISDESELELSESEVVKSEKQ 60  
Db 1 MSDESYQNSTNQIPRSDVLDIRNQITNDCAISDESELELSESEVVKSEKQ 60  
Qy 61 QOHQETSDNAKPLTRKSGSSIKKSNLTKDRITWPMKSLGDDTINSCHKNNYNS 120  
Db 61 QOHQETSDNAKPLTRKSGSSIKKSNLTKDRITWPMKSLGDDTINSCHKNNYNS 120  
Qy 121 SLRKFYLDKNTDNNSTNNHTLAIPPIPIPIITNANKRRKSQLENPLKIKKTI 180  
Db 121 SLRKFYLDKNTDNNSTNNHTLAIPPIPIPIITNANKRRKSQLENPLKIKKTI 180  
Qy 181 GRNNSNFENDLVSPMTKMTNDSEDITNTSTANHKMLGIGATLGVGTGTATATATA 240  
Db 181 GRNNSNFENDLVSPMTKMTNDSEDITNTSTANHKMLGIGATLGVGTGTATATATA 240  
Qy 241 AAGRPRSSIDSEADSHASRSQETEDVCFPMVGDHVRVNGIDFDEIDFIREEREA 300  
Db 241 AAGRPRSSIDSEADSHASRSQETEDVCFPMVGDHVRVNGIDFDEIDFIREEREA 300  
Qy 301 YLOQMIAKNIILRDEPQNLKANTTSGASRHPYHHNNKNNKNGDGGSSMAALKYT 360  
Db 301 YLOQMIAKNIILRDEPQNLKANTTSGASRHPYHHNNKNNKNGDGGSSMAALKYT 360  
Qy 361 PNILKKTLSRFEETHENSSESEIYELTKQQPYKDDQLSLTSTSTSGSGSOVK 420  
Db 361 PNILKKTLSRFEETHENSSESEIYELTKQQPYKDDQLSLTSTSTSGSGSOVK 420  
Qy 421 FGARISDINGGSLPDRFSLPHSESETHADPISLVSPOQSVRLDPNGEETWMLDC 480  
Db 421 FGARISDINGGSLPDRFSLPHSESETHADPISLVSPOQSVRLDPNGEETWMLDC 480  
Qy 481 TCPDSEKMLAKAFGHPHTAEDIRMQETREKVELFKSYVPCFHTFEADKESDYLEP 540  
Db 481 TCPDSEKMLAKAFGHPHTAEDIRMQETREKVELFKSYVPCFHTFEADKESDYLEP 540  
Qy 541 INVYVWFHDLTFHSPISHANRVRRLQRYVDVSADMLCYALIDEITDGFAPVI 600

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.  
OM protein - protein search, using sw model  
Run on: June 13, 2006, 14:20:38 ; Search time 182 Seconds  
(without alignments)  
2346.616 Million cell updates/sec

Title: US-10-018-105A-4  
Perfect score: 4820  
Sequence: 1 MSDESYQNSTNQIPRSDVLDIRNQITNDCAISDESELELSESEVVKSEKQ 922  
Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5  
Searched: 2097797 seqs, 463214858 residues  
Total number of hits satisfying chosen parameters: 2097797  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:  
1: /EMC\_Celerra-SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pcp:  
2: /EMC\_Celerra-SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pcp:  
3: /EMC\_Celerra-SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pcp:  
4: /EMC\_Celerra-SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp:  
5: /EMC\_Celerra-SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp:  
6: /EMC\_Celerra-SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pcp:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4814	99.9	922	4	US-10-032-585-7675 Sequence 7675, Ap
2	1679	34.8	859	2	US-08-945-749-1 Sequence 1, Appli
3	1666	34.6	858	2	US-08-945-749-2 Sequence 2, Appli
4	1099	22.8	663	4	US-10-128-714-3563 Sequence 3563, Ap
5	1099	22.8	663	4	US-10-128-714-8563 Sequence 8563, Ap
6	856	17.8	969	2	US-08-945-749-3 Sequence 3, Appli
7	345.5	7.2	154	3	US-09-864-408A-2266 Sequence 2266, Ap
8	298	6.2	94	4	US-10-425-115-321932 Sequence 321932, Ap
9	234.5	4.9	1198	6	US-08-945-749-8 Sequence 8, Appli
10	234.5	4.9	1198	6	US-11-097-143-7905 Sequence 7905, Ap
11	213.5	4.4	380	2	US-08-945-749-7 Sequence 7, Appli
12	213.5	4.4	386	4	US-10-156-761-9574 Sequence 9574, Ap
13	201	4.2	371	4	US-10-156-761-10644 Sequence 10644, A
14	182.5	3.8	373	3	US-09-738-626-3568 Sequence 3568, Ap
15	179.5	3.7	373	6	US-11-082-389-248 Sequence 248, App
16	179.5	3.7	373	6	US-10-451-467A-574 Sequence 574, App
17	179.5	3.7	2344	4	US-10-732-923-13978 Sequence 13978, A
18	179	3.7	354	4	US-10-156-761-13791 Sequence 13791, A
19	179	3.7	1199	6	US-11-097-143-1614 Sequence 1614, A
20	175.5	3.6	1246	6	US-10-097-143-11433 Sequence 11433, A
21	172.5	3.6	3228	5	US-10-732-923-13547 Sequence 13547, A
22	172	3.6	2165	4	US-10-156-761-14171 Sequence 14171, A
23	170.5	3.5	343	4	US-10-156-761-14171 Sequence 14171, A
24	169	3.5	1468	6	US-11-097-143-22779 Sequence 22779, A
25	168.5	3.5	816	6	US-10-467-555-32 Sequence 32, Appli
26	166.5	3.5	1046	6	US-11-165-819-6 Sequence 6, Appli
27	166.5	3.5	1046	6	US-11-165-819-26 Sequence 26, Appli
28	166.5	3.5	2261	5	US-10-470-048B-60 Sequence 60, Appli
29	166.5	3.5	2283	4	US-10-172-503-4 Sequence 4, Appli
30	166.5	3.5	2283	6	US-11-020-509-4 Sequence 4, Appli
31	166	3.4	688	4	US-10-032-585-7876 Sequence 7876, Ap
32	165	3.4	2112	6	US-11-097-143-8001 Sequence 8001, Ap
33	165	3.4	4498	4	US-10-712-124-68 Sequence 68, Appli



Db 541 INVYLVFHDGLITFHSPIHPANVRVRLQADYVDSADWLCYALIDEITDGPAPI 600  
Qy 601 HGIEYEADEADAVTARDTDFSSMLQIGESRRKVTMLRLLSGKADVIKFAKRCQE 660  
Db 601 HGIEYEADEADAVTARDTDFSSMLQIGESRRKVTMLRLLSGKADVIKFAKRCQE 660  
Qy 661 ANSSGGYQRYNLAQOQVSPNSNKKITEMFSKITLIGTMLVPLNLVTGLFGNVR 720  
Db 661 ANSSGGYQRYNLAQOQVSPNSNKKITEMFSKITLIGTMLVPLNLVTGLFGNVR 720  
Qy 721 NPTGNNTNTNTTSGSPSPPOQOHHGINTKSPIDARPRADIALYLGDIQDHIITMQ 780  
Db 721 NPTGNNTNTNTTSGSPSPPOQOHHGINTKSPIDARPRADIALYLGDIQDHIITMQ 780  
Qy 781 NLAYEKFISRSHSNLAQOQVSPNSNKKITEMFSKITLIGTMLVPLNLVTGLFGNVR 840  
Db 781 NLAYEKFISRSHSNLAQOQVSPNSNKKITEMFSKITLIGTMLVPLNLVTGLFGNVR 840  
Qy 841 VPGEGTNLQWFGVGLVLIPIIIGSFIFAQWMLKLNNSIEQNGNRPFIHSSRSI 900  
Db 841 VPGEGTNLQWFGVGLVLIPIIIGSFIFAQWMLKLNNSIEQNGNRPFIHSSRSI 900  
Qy 901 RSLGLKHGKNSIISPNKYE 922  
Db 901 RSLGLKHGKNSIISPNKYE 922  
RESULT 2  
US-08-945-749-1  
; Sequence 1, Application US/08945749  
; Publication No. US2002013880A1  
; GENERAL INFORMATION:  
; APPLICANT: GARDNER, Richard C  
; APPLICANT: MacDIARMID, Colin W  
; APPLICANT: HAY, Robert J  
; APPLICANT: Auckland Uniservices Limited  
; APPLICANT: New Zealand Pastoral Agriculture Research Institut  
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE  
; FILE REFERENCE: 08/945,749  
; CURRENT APPLICATION NUMBER: US/08/945,749  
; EARLIER FILING DATE: 1998-01-12  
; EARLIER FILING DATE: PCT/NZ96/0035  
; EARLIER FILING DATE: 1996-05-01  
; EARLIER FILING DATE: NZ 272039  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 859  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-08-945-749-1  
Query Match 34.8%; Score 1679; DB 2; Length 859;  
Best Local Similarity 41.5%; Pred. No. 3.5e-107;  
Matches 405; Conservative 137; Mismatches 258; Indels 176; Gaps 31;  
Qy 1 MSSESYYQNTNQPISRDEVLDDHRNQINDCAISDSDELEL---KSESESVK 57  
Db 1 MSSSSS---SSSSPYLSR-----NSLANTMYSMTEDHTGLYDHRQHPDLPVRH 49  
Qy 58 EKQOQHQBITSNAKPLTRKSGSIKKS-----NLTOKRITNPMISLGGDDTINSGH 112  
Db 50 QPPTLKNKSIKATKPSIPKQKATRYNSHVGVSPGRGMDPDEQCGMDSTV-AHH 108  
Qy 113 KNR-----NVMSSL-----RKDFYKNDTNDSTNNHT---HLAIPFIPPTII 156  
Db 109 QURASALLSNARPSRLAHSMHQRLYVESNI-----HTPPKDVGVKRDYTMSSSTA 161  
Qy 157 TNANKSRKSOLENPLPKKTKTIGRNNNNFENDLVSPMTKMTNDSEDIYNT----- 210  
Db 162 SSGNKS-KLSASSASPIKTVR-----KSSLVSPVLBIPEKSDTHSKLAKPK 210  
Qy 211 ---STTANMKLGIGATTIGVGTCTATA---TATAAAGRPSRSSIDSEADSHGRSQ 264  
Db 211 RYVTTSAHSIN-PAVLITKSTQSKSDADDTLERKPYRMNTRASFSDIV-SQGRDSQ 268  
Qy 265 ETEEDVCFPMVGD-HIRVNGIDFIDEFIREEREAYLQKQIAK-NILRIDFQNLK 322

Db 269 ETEEDVCFPMVGPQLTRVNGIDFDELEEYAFANAE---KSQFLASQVNPQKYSNVSQ 325  
Qy 323 -----NNTTSGARHPYHHHNNKNGGGGSGNAALKYTPKINILAKTLRFPETH 376  
Db 326 DIGFTSTSTSGSS-----AALKYTPR--VSQTEKSESTN 359  
Qy 377 ENSSSSEIYELKTKQPPYKYDDQLSLTSSSTSGSGGVKFGG-ARISDINGGSL 435  
Db 360 ET-----EHEKKEDEHEKIKPSLHPGISGKNKVEGEENENIPSDPAYCVSQGTDFQI 414  
Qy 436 PDRFSLFSESETHAPDIPSLVSPQSVROLFRNGEETWMLDCTPTDSEMKLAKAF 495  
Db 415 PNRSPFCSESDETVTHASDIPSLVSEGOFTFELFRGGEPTWMLDSCPTDDMRCAKAF 474  
Qy 496 GHPUTAEDIRMOETREKVELFKSYFVCFHTFPAKRESESDYLEPINVIYVPHDGLTF 555  
Db 475 GHPUTAEDIRMOETREKVELFKSYFVCFHTFPAKRESESDYLEPINVIYVCRSGVLT 534  
Qy 556 HFSPIHPANVRVRLQADYVDSADWLCYALIDEITDGPAPVHIGIEYEADEADAVP 615  
Db 535 HFGPISHCANVRVRLQADYVDSADWLCYALIDDTDSFAPVIOIEYEADEADAVP 594  
Qy 616 TARDTDFSSMLQIGESRRKVTMLRLLSGKADVIKFAKRCQEANSSGGYQRYNLAQ 675  
Db 595 MARMDPFAMLQIGESRRKVTMLRLLSGKADVIKFAKRCQDEANGI----- 643  
Qy 676 OQOQOAPPPPPNPIITSPINSLNLSLGTSTGGVGVGGINGFNPNPTGNNTNTNTTGT 735  
Db 644 -----GPAITSQIN-IANLQARQNA-----SHKNNSTTVPNY 678  
Qy 736 SPSPPOQOQHHGINTKSPIDARPRADIALYLGDIQDHIITMQLNLAVERIFSRSHN 795  
Db 679 APTTSQ-----PRGDIALYLGDIQDHIITMQLNLAVERIFSRSHN 720  
Qy 796 YLAQLQVESFNSNKKITEMFSKITLIGTMLVPLNLVTGLFGNVRVPGEGTNLQWFGI 855  
Db 721 YLAQLQVESFNSNKKITEMFSKITLIGTMLVPLNLVTGLFGNVRVPGEGTNLQWFGI 779  
Qy 856 VGVLIPIIIGSFIFAQWMLKLNNSI---EQNGNRPFIHSSRSIRSLGLK 907  
Db 780 LGVLLLVGLVGLFASVYKIRIDPPATLNAEBSAKSVISSFLPKRKNRPNDRSKNIN 839  
Qy 908 HGG-NKSIISPNKY 921  
Db 840 RAGPSNKSVALSPSY 855  
RESULT 3  
US-08-945-749-2  
; Sequence 2, Application US/08945749  
; Publication No. US2002013880A1  
; GENERAL INFORMATION:  
; APPLICANT: GARDNER, Richard C  
; APPLICANT: MacDIARMID, Colin W  
; APPLICANT: HAY, Robert J  
; APPLICANT: Auckland Uniservices Limited  
; APPLICANT: New Zealand Pastoral Agriculture Research Institut  
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE  
; FILE REFERENCE: 08/945,749  
; CURRENT APPLICATION NUMBER: US/08/945,749  
; EARLIER FILING DATE: 1998-01-12  
; EARLIER FILING DATE: PCT/NZ96/0035  
; EARLIER FILING DATE: 1996-05-01  
; EARLIER FILING DATE: NZ 272039  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 858  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-08-945-749-2  
Query Match 34.6%; Score 1666; DB 2; Length 858;  
Best Local Similarity 40.9%; Pred. No. 2.8e-106;  
Matches 401; Conservative 136; Mismatches 221; Indels 233; Gaps 35;  
Qy 39 DSEDELEKSESEFV-----KSEKQOQHQBITSNAKPLTRKSGSIKKSNTLTKD 93

[illegible]

305 -----IVEKLQNEENETRFQFSSSESQSTVHAAELGDLVLPGD 342
464 SVRLDFRNGEE--TWLDCCTPTDSEKMLAKAFGIHPLTAEDIRMOETREKVELFKSYT 521
343 TRDLFQLGPEGVWMLDVLNPTTESEVAALSRAPSIHPLATEDILTQEAEREKVELFKQY 402
522 FVCFHTE--EADKESDYLEPINVYVVFHOGIUTPHSPISHPANRRVRQRQYDVDS 580
403 FVCFHTEFVQLDKTDREPMFVNFYVVRDGVLSFSPTEPHAAVNRKRIKLRDVSLS 462
581 ADMLCALIDEITDGFAPVHIGIEYADAEADVFTARDTDFSSMLQRIGESRRKVTLM 640
463 SDWICYAMIDDIIVDSFGVIRETEIESEAEDELVPIARVDDFESFLPRIGLKKVNSLM 522
641 RLSCGKADVIKFAKCOEANSSSGYQRYQVNLQOOQOQAPPPPPNPIITSPINSLNL 700
523 RLSCGKADVIRGFSRCNE-----541
701 NSLGTSTGGGVGGVGNFPGNPTGNNTNTNTTGTSPSPQOQOQGGITNKSPIPDARP 760
542 -----QVSVT-----P 547
761 RADIALYLGDIQDHIITMFLAYEKIFSRSHSNYLAQIQVESFNSNNKITEMFSKITL 820
548 RGDIGLYGDIQDHRVVTMSNLAFKMLSRSHNYLAQLNVNLVNLGNHVKVLSKVTL 607
821 IGTMLVPLNLVLTGLFGMNVVRPQEGTNLGMFFGIVGVLIIFIIGSFIPAQMW 873
608 IATMLVPMNLICGLFGMNVVRPQEGTGLGMFFGIVGVIAAVIVLSGIAARYY 660
RESULT 6
US-08-945-749-3
; Sequence 3, Application US/08945749
; Publication No. US2002013880A1
; GENERAL INFORMATION:
; APPLICANT: GARDNER, Richard C
; APPLICANT: MacDIARMID, Colin W
; APPLICANT: HAY, Robert J
; APPLICANT: Auckland Uniservices Limited
; APPLICANT: New Zealand Pastoral Agriculture Research Institut
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE
; FILE REFERENCE: 08/945.749
; CURRENT APPLICATION NUMBER: US/08/945.749
; EARLIER FILING DATE: 1998-01-12
; EARLIER APPLICATION NUMBER: PCT/NZ96/0035
; EARLIER FILING DATE: 1996-05-01
; EARLIER APPLICATION NUMBER: NZ 272039
; EARLIER FILING DATE: 1995-05-01
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-08-945-749-3
Query Match 17.88; Score 856; DB 2; Length 969;
Best Local Similarity 24.18; Pred. No. 4e-50;
Matches 270; Conservative 143; Mismatches 248; Indels 460; Gaps 29;
QY 11 STTNQIPRSDEVLDDHNRQITNDCAISDEDELKSESEVVKSE-----58
DB 44 SDSRRPTQLLDHNLQHNHGOITDFQI--DSWGMHESDSTNDIKSEDPKLGAFIDHR 102
QY 59 -----KQOQHQBITS--NAPLTRKSGSSIKKKSNLTKDKRITNPM--LSGGDDTINSCH 112
DB 103 PSMQOPREGQSVSTVQPIPMKFTSPYKPKACLRPSQNRSLVSDLSPSE--LESML 160
QY 113 KRNYNMSSLRKDFYLDN--TDDNSTNNHTHLLAIPITPIPTIITNA-----NKSRRK 165
DB 161 KRR-----KSVHKSFPVDNSPTDRQSNANNVDVVDV--ALMHNHVNNASTGVNDNSKRRK 215
QY 166 SQLENLPLIKKTKTGRNNSNPNFENDLVSPMTKMTWDSSEDTNTSTANRKLIGIGATT 225
DB 216 -----KKR--GSDSSN--KN--SKSTSSDSNDEEDSYN-----244
QY 226 LGVGTGTATATATAAAGRRPSRSDSIDSEADSHASRSSQFTBEDVCFPM--VGDHVRNG 283

761 RADIALYLGDIQDHIITMFLAYEKIFSRSHSNYLAQIQVESFNSNNKITEMFSKITL 820
548 RGDIGLYGDIQDHRVVTMSNLAFKMLSRSHNYLAQLNVNLVNLGNHVKVLSKVTL 607
821 IGTMLVPLNLVLTGLFGMNVVRPQEGTNLGMFFGIVGVLIIFIIGSFIPAQMW 873
608 IATMLVPMNLICGLFGMNVVRPQEGTGLGMFFGIVGVIAAVIVLSGIAARYY 660
RESULT 5
US-10-128-714-8563
; Sequence 8563, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus
and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128.714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8563
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8563
Query Match 22.88; Score 1099; DB 4; Length 663;
Best Local Similarity 32.78; Pred. No. 3.5e-67;
Matches 292; Conservative 98; Mismatches 241; Indels 262; Gaps 22;
QY 4 SESVYQNSTTNQIPRSDEVLDDHNRQITNDCAISDEDELKSESEVVKSEKQOQH 63
DB 7 SEKYVCLTRESTVPE--LDDHRFQI-----DSPPRIEATADI--SLSRQNTAQHAY 54
QY 64 HQEITSDNAKPLTRKSGSSIKKKSNLTKDKRITNPMSLGCGDDYTNSGHKNRYNMSSLR 123
DB 55 HQETPQ-----RPDLLSIQDALREAGSLSRDFEQALD 87
QY 124 KDFYLRD-----NTDNDSTNNH-----THLAIPITPIPTIITNANKSRK--SQ 167
DB 88 DRSKGDINTLGRPSVDPNGNVHGRHWTSRTHQEL-----ANMSRESPPSA 134
QY 168 LENLPLIKKTKTGRNNSNPNFENDLVSPMTKMTWDSSEDTNTSTANRKLIGIGATTLG 227
DB 135 RESPPP-----NSVEAPADPRRERANTLE-----SHAAPDLAQLQRTVSG 176
QY 228 VGTGT--TATATATAAAGRPSRSSIDSEADSHASRSSQETEDVCFPMGVDHVRNGIDF 286
DB 177 ---GTHPRRPFTFSNASAIRPQDIOLEPN-----DESCVPTQEOPGRIPVIDY 222
QY 287 DBIDFTREEREAYLQKQIAKNILRIDEFQNLKSNKNTTSGASRHP---YHHSSNNKK 343
DB 223 EBLEPFA-----LSRQMKPSTR--R-----RQSLSSQSGRPVFPVYDLRPGLRKS 266
QY 344 NNGGDDGSSMAALKYTPKNILKTLKTRFEFTHENSSSSSEETVELKTKQPPKYDDQLS 403
DB 267 DVEGEKASSA-----DRSSDLMADLKTAKDYVANYVDEKD 304
QY 404 LTSSTSTSGSGSQVFGGARISDINGGSLPDRFSLPHSESEBETHAPDIPSLVSPQ 463

US-09-864-408A-2266

Query Match  
Best Local Similarity 7.2%; Score 345.5; DB 3; Length 154;  
Matches 79; Conservative 20; Mismatches 39; Indels 39; Gaps 6;

Qy 351 GSSMAALKYTPKNILKTLRPFETHSSSSSEIEYELKTKQPPYKDDQLSLTSSTSS 410  
Db 1 GSS-AALKYTPKE-----ISR-----TLEKNCVNEYVSENNES--VREDDKPD----- 42

Qy 411 TSGSGGVVFGGARGISDINGG-----SLDRFSLPHSESEETIHAPD 454  
Db 43 -----HPDVTFRNKIRGEKGNDSYSRAVYTLQNTVEYQIPRSFSESTVHAASD 97

Qy 455 IFSLSVFGQVRLDPRNGEETWMLDCTPTDSMKMLAKAFGIHPLTAEDIRMOETR 511  
Db 98 IFSLSISGQTFVELFKGSDPTWMLDSCPTDDMRCTAKTFTGIHPLTAEDIRMOETR 154

RESULT 8  
US-10-425-115-321932  
; Sequence 321932, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222) B  
; CURRENT APPLICATION NUMBER: US/10/425.115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369126  
; SEQ ID NO 321932  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MNT4577\_56667C.1.pep  
US-10-425-115-321932

Query Match  
Best Local Similarity 6.2%; Score 298; DB 4; Length 94;  
Matches 54; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

Qy 573 LRDYVDSADMLCYALIDETDGPAPVHIGIEYADAEIDAVTARTDPTFSMLQRIGES 632  
Db 1 LRDYVSSDMCYAMIDDDIVDSFGPIMRDVIIESEADIEDHVFARVEDFGTFLPRIGDL 60

Qy 633 RKVNTLMRLSLGKADVIKFAKQEE 660  
Db 61 RKVNTLMRLSLGKADVIKFAKQEE 88

RESULT 9  
US-08-945-749-8  
; Sequence 8, Application US/08945749  
; Publication No. US2002013880A1  
; GENERAL INFORMATION:  
; APPLICANT: GARDNER, Richard C  
; APPLICANT: MacDIARMID, Colin W  
; APPLICANT: HAY, Robert J  
; APPLICANT: Auckland Uniservices Limited  
; APPLICANT: New Zealand Pastoral Agriculture Research Institut  
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE  
; FILE REFERENCE: 08/945.749  
; CURRENT APPLICATION NUMBER: US/08/945.749  
; CURRENT FILING DATE: 1998-01-12  
; EARLIER APPLICATION NUMBER: PCT/NZ96/0035  
; EARLIER FILING DATE: 1996-05-01  
; EARLIER APPLICATION NUMBER: NZ 272039  
; EARLIER FILING DATE: 1995-05-01  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Synechocystis sp.

US-09-864-408A-2266

Query Match  
Best Local Similarity 44.6%; Pred. No. 6.8e-16;  
Matches 79; Conservative 20; Mismatches 39; Indels 39; Gaps 6;

Qy 351 GSSMAALKYTPKNILKTLRPFETHSSSSSEIEYELKTKQPPYKDDQLSLTSSTSS 410  
Db 1 GSS-AALKYTPKE-----ISR-----TLEKNCVNEYVSENNES--VREDDKPD----- 42

Qy 411 TSGSGGVVFGGARGISDINGG-----SLDRFSLPHSESEETIHAPD 454  
Db 43 -----HPDVTFRNKIRGEKGNDSYSRAVYTLQNTVEYQIPRSFSESTVHAASD 97

Qy 455 IFSLSVFGQVRLDPRNGEETWMLDCTPTDSMKMLAKAFGIHPLTAEDIRMOETR 511  
Db 98 IFSLSISGQTFVELFKGSDPTWMLDSCPTDDMRCTAKTFTGIHPLTAEDIRMOETR 154

RESULT 8  
US-10-425-115-321932  
; Sequence 321932, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222) B  
; CURRENT APPLICATION NUMBER: US/10/425.115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369126  
; SEQ ID NO 321932  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MNT4577\_56667C.1.pep  
US-10-425-115-321932

Query Match  
Best Local Similarity 6.2%; Score 298; DB 4; Length 94;  
Matches 54; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

Qy 573 LRDYVDSADMLCYALIDETDGPAPVHIGIEYADAEIDAVTARTDPTFSMLQRIGES 632  
Db 1 LRDYVSSDMCYAMIDDDIVDSFGPIMRDVIIESEADIEDHVFARVEDFGTFLPRIGDL 60

Qy 633 RKVNTLMRLSLGKADVIKFAKQEE 660  
Db 61 RKVNTLMRLSLGKADVIKFAKQEE 88

RESULT 9  
US-08-945-749-8  
; Sequence 8, Application US/08945749  
; Publication No. US2002013880A1  
; GENERAL INFORMATION:  
; APPLICANT: GARDNER, Richard C  
; APPLICANT: MacDIARMID, Colin W  
; APPLICANT: HAY, Robert J  
; APPLICANT: Auckland Uniservices Limited  
; APPLICANT: New Zealand Pastoral Agriculture Research Institut  
; TITLE OF INVENTION: ALUMINUM RESISTANCE GENE  
; FILE REFERENCE: 08/945.749  
; CURRENT APPLICATION NUMBER: US/08/945.749  
; CURRENT FILING DATE: 1998-01-12  
; EARLIER APPLICATION NUMBER: PCT/NZ96/0035  
; EARLIER FILING DATE: 1996-05-01  
; EARLIER APPLICATION NUMBER: NZ 272039  
; EARLIER FILING DATE: 1995-05-01  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 387  
; TYPE: PRT  
; ORGANISM: Synechocystis sp.

Query Match	6.2%	Score 298;	DB 2;	Length 387;
Best Local Similarity	23.0%;	Prod No. 4.8e-12;		
Matches 108;	Conservative 70;	Mismatches 143;	Indels 148;	Gaps 18;

  

QY	437	DRESLPH-----SESEETHADIPS-----LV--SQOSVR--DLFNGEEFW-----W 477
DB	27	DYDYTFDEPGSEPGTSLIEDPPSRIVUDYVPSHVARSDISPNALPYGLTGVTSW 86
QY	478	LDCT-CPTDSMKMLAKFGTHLTAIDRDMQETREKVELFSYVYVCHFTPEADVEDSD 536
DB	87	MDTEGLSGSEVLKEGVEIFKLLLEDLVVQPAKVEDYNDVAVTAHVRNRE--BD 145
QY	537	VLEPIVW-----YVVFHGILITPSPISHPAMRRVRQLRDYY--DVSAWLCYAL 588
DB	146	GFSEQGVFVGLKRYLLTQFGHILDCNPL-----RRRTNQGVQCGQADYLCYL 199
QY	589	IDELITGPAVHIGIYEADIAIEDVATRTDPSNQRIGESRKVNTWMLLSGKAD 648
DB	200	IDMLDTEYPLLEDYETREALEDTII--RPNP--SSLMEEYHTRRELLARLL----- 251
QY	649	VIKNFARKQBSEANSSGGYQRYNLQOQOQOQAPPPPNPIYSPINSLTMSLGTSG 708
DB	252	-----WPLRRVWMLLRDT----- 265
QY	709	GGVGVGGINFGNPCTNNTNTNTNTTSGSPPOQOQGHITNKSFPIDAPRADTALY 768
DB	266	-----TNSIV-----ADRVRYF 278
QY	769	GDIQDHITFQNLAYEKIESRSHSNYLAQOVESNNKNITEMFSKITLIGTMLVLP 828
DB	279	RDCYDHIIQVLVDIEAYRELASSIMEVYNTAM-----SNKNEVAKFLVISTIFIPL 331
QY	829	NLVTVGLGHNVVRVPGGGTNGLWFFGIVGLVIFIITIGSFIAQW---WL 874
DB	332	TFIAGYVGNFNKNP--LNSRGGYVYIMVLLIAGSGLYFFWRKGLM 378

Query Match 4.9%; Score 234.5; DB 6; Length 1198;  
Best Local Similarity 19.7%; Pred. No. 5.7e-07;  
Matches 185; Conservative 114; Mismatches 291; Indels 349; Gaps 41;

QY 57 SEKOOHQHQRJTSNAPUTRSGSS-----TKKS-----87

Db 37 SHOOOQQQLLPHHHHDQWLAAGSSPLFPFYSHLQLOQKDATATGPAAAAARAAEAATT 96

QY 88 -----NLTKDRIITNPWSLGCGDT---INSGHKRNVM-----SSLRKDFVLKDN 131

Db 97 SANADNFSSLOTIDASQOGGISLGLCDPRPVPASPNPHSNWMLTGTATAATTTTNN 156

QY 132 TDDNSTNNHTHLAIPPIPT-----PILTANKSRKSOLENLPPLIKKT 179

Db 157 NNNNTNNNNNNVNAEATVRPSNGNSVIIESTWSPANILPTHRSENICDPALQKN 216

QY 180 IGRUNSNFTENDLVSP-----MTXMKTNDESDTW-----NSTYAN--RMGLGTG-----222

Db 217 PQNPNGNN--SSIIVPVVEYHQLKPLEVNSSTSVTSNPLSSTTAQLLDPEVOVGKDDCHI 275

QY 223 -ATTLLGVGTG-----TATATATAAGR--RPSRSSI-----251

Db 276 STTTTTPGSGSGSAGSGSGSGSIARTIGTATPTTTTMSNTANTPTRASSLIHSIELA 315

QY 252 -----DSBADSHARSQSOTPEDVCPPKVGDHVRVN-----GIDPD 287

Db 336 ASSCA PRAASPNSNHTSSATTPQOQQOQHMGSGNHSGSNLSDDBSMSEDEFGLEID 395

QY 288 EIDFEIRERREAYLOKMIKNILRIDFO-----NLSKNTTSG-----ASHR 332

Db 396 DNGEMDQQATVQFISSLNISPPSQMLEQQQSPSPALAAAGCNSSNNAASGSNNNASGN 455

QY 333 PYVHHNSNNKKNN-----GGGGSGSMAALKTY-----360

Db 456 NTSSSSNNNNNNNDANDHVLTKEPEHYNAVYTLQLAGGGSGSGNQHHNSHNHGHHQ 515

QY 361 -----PKNILKLTLSRFETHNSSSSSEIEELKTKQOPPY 396

Db 516 QOQQOQQOQQOQHQQOQEHYQOQQOQNANNA--NQF-----NSSSYSVIYNPDSQYIPT 569

QY 397 KYDOLSLTSSSTSGSGSGSGVKFGARISDCINGGS-----LPDRFLSFHESEE 448

Db 570 GYQB--TTSSHQSQSGGGGG--GGGNLLNGSGSGSAGGYMLLP-----611

QY 449 TTHAPDIPSLSVQSVVDLPFBGBETMWLDCCTPTOSEKMLAKAFGI-----497

Db 612 -----QAASSSGNGNN-----PNAGHMSSGVSNGSGGAGGAGGN 648

QY 498 -----HPLTAEDIRMOETREKV--LFKSYYPVCFPTHFAEDKESDYLEPINVYTVFPHD 350

Db 649 SGCNPNMGTSATPGCHGEVIDFKHLFEELPCVC-----GDKVSG-----YHY 691

QY 551 GLTFHFPSIHANVRVRVRLRDYDVYSAWLCYALIDEITDGFAPVTH-----GI 603

Db 692 GLTACE-----CKGFPPKTVQMKVYTCV--AESRCH--IDKQKRCPYCRFOKCLEVM 744

QY 604 EYEADATEDAVPTARDTDFSSMLOR-----IGESRRKVTMLRLLSKADVIMKFAKRCQ 658

Db 745 KLEA-VRAARMRGGRN-KFQPMYTRDRARKLQWRQOLAOLALNSNGSPDIK-----795

QY 659 BEANSSSGYQRVNLQOQQO-----QAPPPPNNPITTSINSTLNLSLGTG 708

Db 796 -PTPIPSGYQQAYPENNNIQEIQIPOVSSITQSPDSSPSPI-----AIALGOVNASTG 847

QY 709 -----GGVGVGGIFGCPNPTGNNTNTNTNTNTTGS 736

Db 848 GVIAEPNAGTGGSGGGLN-CPSSVNGNNGSSNGSNGN 885

RESULT 11  
US-08-945-749-7  
; Sequence 7, Application US/08945749  
; Publication No. US2002013880A1  
; GENERAL INFORMATION:  
; APPLICANT: GARDNER, Richard C  
; APPLICANT: MacDIARMID, Colin W  
; APPLICANT: HAY, Robert J  
; APPLICANT: Auckland Uniservices Limited  
; TITLE OF INVENTION: New Zealand Pastoral Agriculture Research Institut  
; FILE REFERENCE: 08/945,749  
; CURRENT APPLICATION NUMBER: US/08/945,749

```

1 CURRENT PILLING DATE: 1998-01-12
2 EARLIER APPLICATION NUMBER: PCT/NZ96/0035
3 EARLIER PILLING DATE: 1996-05-01
4 EARLIER APPLICATION NUMBER: NZ 272039
5 EARLIER PILLING DATE: 1995-05-01
6 EARLIER PILLING DATE: 1995-05-01
7 NUMBER OF SEQ ID NOS: 16
8 SOFTWARE: Patent In Ver. 2.0
9 SEQ ID NO 7
10 LENGTH: 360
11 TYPE: PRT
12 ORGANISM: Synechocystis sp.
13 US-08-945-749-7

```

Query Match 4.4%; Score 214; DB 2; Length 380;  
Best Local Similarity 19.2%; Pred. No. 3.1e-06;  
Matches 84; Conservative 72; Mismatches 141; Indels 140; Gaps 14;

478	LCCTCPDSE-----NEMLAKEGHLPTAEDTBMQETREKVLKFSYVP	522
QY		
DB		
59	LDCEAVUTSVESWINIDCLGNHTWEQLGEVKLHPVALBEDIUNVPQPKVEYENHLI	118
QY		
DB		
523	VCPTHFEADKESDYLEPIWYIVWHDGILUTHFSP--ISHPANRRVRVROLRDYV-DVS	580
QY		
DB		
119	PISRWMTDQSQTSFISQISFGLGH-YLUTHQEPEKYDCLSVREIRKTKGAIKQKN	177
QY		
DB		
581	ADM/CYALIBDITGPAPVHNGIEYEADAIEDAVFTARDTDFSNLQRIGSSRRKVMUTLM	640
QY		
DB		
178	ADYLFYALIDAIIDGFPV---MEVYGLBVQSQSEIISCTPNKSLAKIHQLQODULLNR	234
QY		
DB		
641	RLUSGKADVIMKPAKRCQEEANSQSYQRYNLQOQQQQAPPPPNPPIITSPINSTLNL	700
QY		
DB		
235	RATWQPDAL-----	244
QY		
DB		
701	NSLGTSTGGVGVGGINFGNPNTNNNTNTNTGTSPFPQQQOHHGHTKSPFIDARP	760
QY		
DB		
245	NSL-----LRDGSGLDSD-----	257
QY		
DB		
761	RADIALYLGDIQDIHTWFQNLAYEKIFSRHSNYLQAQVSEFNSNNKITEMFKITPL	820
QY		
DB		
258	--EVRVLRDCYDHTIQLDMDIETRYDLASNLDTIYLSVV-----SNRNWETMKTLTV	308
QY		
DB		
821	IGTMLVPLNLVTGLQMNVRVPCEGQTN---LQWFFEG--IVGVLIFIIIGSFIAQW---	872
QY		
DB		
309	ISSIFETPLTITAGIYGNMNFN-PDKSPWNPNELWYNGYGVPIWVMTLGVGNALYFFWRKG	367
QY		
DB		
873	WLKLNNSIEQONNGNR	889
QY		
DB		
368	WFRNL-----VEKNGR	380
QY		
DB		

```

RESULT 12
US-10-156-761-9574
; Sequence 9574, Application US/10156761
; Publication NO. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: YAMAGUCHI, MASAHITO
TITLE INVENTOR: NOVEL POLYNUCLEOTIDES
FILE REFERENCE 249-262
CURRENT APPLICATION NUMBER: US/10156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9574
LENGTH: 386
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9574

```

Query Match	4.4%	Score 213.5;	DB 4;	Length 386;
Best Local Similarity	19.6%	Pred. No. 3.4e-06;		

	Matches	86; Conservative	67; Mismatches	150; Indels	135; Gaps	13;
Qy	459	VSPQQVVDLFRNGE----	TWMLDCTPDSSEKKLKAFGHPHUTADIRMQETREKV	514		
Db	62	VSPASCADTFRELDREFSGMAGILAPRTESSELLABEDFLUPLAVEDAMEHQRPKL	121			
Qy	515	ELPKSYVVCWHT--PEADRESDEYLPIINVIYVPFHGILITPHFSPISHPAKVRVRVQ	572			
Db	122	ERGETLVVLUAARLYDAPEVDFFE---	LHVFGPPGVITYRHGAAPDLUSAVRRHEE	178		
Qy	573	LRDYVDSADWLCVALIDETDFGPAPHVIEVEADAIEDAVTATARDTFSSMLQRIGES	632			
Db	179	TPBLLKGLGEAVLYATLDAAVVGYSVPVAGVQNDIDETEETVFRGD----	PAYSRRRIAYEL	234		
Qy	633	RRKVMTLARLLSGKADVIKMPAKRCQEBSNSSGGYYQRYQLAQOQQOQAPPPIPNDIITS	692			
Db	235	SREMVEFOR-----	SLMAAGFAKYGTD-----	246		
Qy	693	PINSTLNLSNLSGTSGGGVGGVGNFPNPNTGNTWTNTTSTGSPSQOQQOHGITKS	752			
Db	247	PLVGMHL-----	SLMAAGFAKYGTD-----	265		
Qy	753	FPDPARPRADIALLYGIQHFI-----	ITMFQNLLA-----	801		
Db	266	-----	BELQRLRVADAVRHTTSERVDFGRFQAQLADILTYNATLVTOQQAENRALA	316		
Qy	802	VESFNSENKIKEMFSKITLGTLVPLNLVTGLFGNVVRPGEGTNLCGWFFG----	IYGV	858		
Db	317	EAGEFEQNSEI-----	KKISSMAAILFAPTLVGTIYGNF-----	EHMPELGSFGYPFAIGL	368	
Qy	859	LIPIIIGSFIP--AQWML	874			
Db	369	MGLVCCVSYIVFKRRGWL	386			

```

RESULT 13
US-10-156-761-10644
; Sequence 10644, Application US/10156761
; Publication NO. US2003011918A1
; GENERAL INFORMATION
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10644
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10644

```

Query Match	4.2%	Score 201;	DB 4;	Length 371;
Best Local Similarity	19.7%;	Pred. No. 2.3e-05;		
Matches 91; Conservative 64; Mismatches 160; Indels 148; Gaps 13;				
QY	436	PDRSLFHSSE-ETIHAPDPSLV-----SFGOSVDRFLRNGEFTW	476	
		:		
DB	13	PSRSMKRGDSAYDTTRDPSTPEAVVDCAIYRDGARVATRPITPHEANVRQRGGPV-	71	
QY	477	WLDCCTPTDSEMKLAKAFGIHPLTAEDIRMQRETKVELFKSYTFVCF---HTPEDK-	532	
		:		
DB	72	WIGLHEPTAEAFAGIFGLHLFAVEDAVQAORPKLYRYDDSLFTVFKTTHYYEHQDL	131	
		:		
QY	533	-SESDYLEPNVIVVFHDGILTFHFSPISHPANVRRRVRLQRLDVVDASDWLCVALIDE	591	
		:		
DB	132	TANSSEWTEGVNCTGRDFPIFVHRHGQGSLRAURLHQDLLQPPELLAKGPSVLHATAIH	191	
		:		
QY	592	ITDGAPVVGIEYEADAIEDAVF-----TARDTDFSSMLORICESRRVMTLRLLSG	645	
		:		

Db 239 -----IIDM-----QHTTSSVTE 251  
Qy 740 PQQOQHGINKSPFPIDARPRADIALYLGDIQDHIITFQNLAYEKIESR---SHSNY 796  
Db 252 VVQR-----LNKDF--IRSGMSEELRAYLDDVADHLTRDNTRVSEYRESLSQILNVAATL 304  
Qy 797 LAQLOQSEFNSNNKKTITLIGTLMVPLNVLVGLFCMVRVPGEGGTNLGWFPG-- 854  
Db 305 VAQR-----NEDMKKISGWAIIIFAPTLVSSYIGMNFIDIMP-----LHMAFGYP 350  
Qy 855 --IVGVLIPIIIGSFIF--AQW 872  
Db 351 LALLAMLGPTLLLYWIFKRSKW 372

RESULT 15  
US-11-082-389-248  
; Sequence 248, Application US/11082389  
; Publication No. US20050244935A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauser, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
; TITLE OF INVENTION: TRANSPORT  
; FILE REFERENCE: BGI-131PCPN  
; CURRENT APPLICATION NUMBER: US/11/082.389  
; CURRENT FILING DATE: 2005-03-16  
; PRIOR APPLICATION NUMBER: US 09/603024  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: US 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 60/143262  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: US 60/151281  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: DE 19930487.4  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: DE 19930489.0  
; PRIOR FILING DATE: 1999-07-01  
; PRIOR APPLICATION NUMBER: DE 19931549.3  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931550.7  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19932134.5  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19941379.7  
; PRIOR FILING DATE: 1999-08-31  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 446  
; SEQ ID NO 248  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-11-082-389-248

Query Match 3.78; Score 179.5; DB 6; Length 373;  
Best Local Similarity 19.9%; Pred. No. 0.00073;  
Matches 100; Conservative 68; Mismatches 145; Indels 189; Gaps 20;  
Qy 422 GGAISDINGCSLPDRFSLPHSESEETIHAPDIPSLVSP-----CQSVR 466  
Db 9 GAIRRRDMLRRYLFD-----SANSTVPPEVSELPTRYVTDGIPKRPPLGATVA 57  
Qy 467 D--LFRNGSETWLDCCTPDSE--MMLAKAFGHPHTAEDIRMQETREKVELFKSYYP 522  
Db 58 DGLKFAEGASNNRMVNSLYPAPSKPAIEELAEAWDLHPTIVEDLLGQQRPKLDREYDIIF 117  
Qy 523 VCFHT--FEADKESDYLEPINVIVVFH-----DGLTFHFSPISH 562  
Db 118 IAIKSAFYDSREVDSE-----FILMKPQAIALCODNQWIDGTSAAFS--SN 166  
Qy 563 PANVRRVRQL---RDYVDVSADMLCYALIDEITDGFAPVHGIEYEAIDAEDAVFTARD 619  
Db 167 PEEDIKRITLLADAELLSSGPRAAAYRLDAIVDGFSPVLGTAIDQEQIERQVP----- 222

Db 192 VVDGYIADAVQDDIDEVETVSPGRKGTGRGTAG-----RIYQLKREVLEFKRAVSP 247  
Qy 646 KADVIKFAKQCEANSSGYQRYQNLQOQQOQAPPPPPPIITSPINSLNLSLGT 705  
Db 248 LRPMLLSER-----PWRLLD-----264  
Qy 706 STGGGVGGGNGFPNPTGNNTNTNTWTTSPPSQOQOQHGINKSPFPIDARPRADIA 765  
Db 265 -----PDIQ 268  
Qy 766 LYLGDIDQDHIITFQNLAYEKIESRSHSNLYLAQLOVESFNSNNKKTITLIGTMI 825  
Db 269 KYFRDVAHLARVQVLGDFDLNLSLQANLAQASVAQNMEDMKRITS-WAAIIAVPTM- 326  
Qy 826 VPLNVLVGLFCMVRVPGEGGTNLGWFPG--IVGVLIPIIIG 865  
Db 327 -----VCGVYGMNF-----EHMPELHMRFGYPVINSVTVLICLG 360

RESULT 14  
US-09-738-626-3568  
; Sequence 3568, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKAZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738.626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3568  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3568

Query Match 3.8%; Score 182.5; DB 3; Length 373;  
Best Local Similarity 19.9%; Pred. No. 0.00045;  
Matches 100; Conservative 69; Mismatches 144; Indels 189; Gaps 20;  
Qy 422 GGAISDINGCSLPDRFSLPHSESEETIHAPDIPSLVSP-----CQSVR 466  
Db 9 GAIRRRDMLRRYLFD-----SANSTVPPEVSELPTRYVTDGIPKRPPLGATVA 57  
Qy 467 D--LFRNGSETWLDCCTPDSE--MMLAKAFGHPHTAEDIRMQETREKVELFKSYYP 522  
Db 58 DGLKFAEGASNNRMVNSLYPAPSKPAIEELAEAWDLHPTIVEDLLGQQRPKLDREYDIIF 117  
Qy 523 VCFHT--FEADKESDYLEPINVIVVFH-----DGLTFHFSPISH 562  
Db 118 IAIKSAFYDSREVDSE-----FILMKPQAIALCODNQWIDGTSAAFS--SN 166  
Qy 563 PANVRRVRQL---RDYVDVSADMLCYALIDEITDGFAPVHGIEYEAIDAEDAVFTARD 619  
Db 167 PEEDIKRITLLADAELLSSGPRAAAYRLDAIVDGFSPVLGTAIDQEQIERQVP----- 222  
Qy 620 TDFSEMLQRIGSRKRVMTLMLLSGKADVIKMFARQCEANSSGYQRYQNLQOQQO 679  
Db 223 -----SGDAVA-----ERYNLQS8-- 238  
Qy 680 QAPPPPPPIITSPINSLNLSLGTSTGGGVGGGNGFPNPTGNNTNTNTWTTSPPSP 739

QY	620	TDFSMQRI	GESREK	VNTLMRL	LSGKAD	VIKMFAKRC	QEBEANS	SSGGYQ	QYQY	NLQOQOQ	679
Db	223	-----	-----	-----	-----	-----	-----	-----	-----	-----	238
QY	680	QAPPPPP	PIITS	PINST	LNLSL	GTSTGG	VGVGIN	FGPNPT	CGNNT	NTNTNT	739
Db	239	-----	-----	-----	-----	-----	-----	-----	-----	-----	251
QY	740	PQOQOQH	ITNKS	FPIDAR	PRADIA	LYLQDI	QDIHI	ITMFQ	NLLAYE	KIFSR	796
Db	252	VVOR	-----	-----	-----	-----	-----	-----	-----	-----	304
QY	797	LAOLQV	SPNSN	KITEM	FSKITL	IGTML	VPLNL	VTGL	PCMN	VRVPC	854
Db	305	VAORQ	-----	-----	-----	-----	-----	-----	-----	-----	350
QY	855	--IVG	VLIFI	IIGSF	IP--AOW	872					
Db	351	LALLAM	LGTLL	LYWIF	RSKW	372					

Search completed: June 13, 2006, 14:23:59  
Job time : 187 secs



OM protein - protein search, using sw model

Run on: June 13, 2006, 13:59:19 ; Search time 200 Seconds  
(without alignments)  
2107.768 Million cell updates/sec

Title: US-10-018-105A-4

Perfect score: 4820

Sequence: 1 MSDESRYQNTNQPIPRP.....LGLKKHGKNSIIFPNKYE 922

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*  
10: Geneseq2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4820	100.0	922	3 AAB36515	Aab36515 Candida a
2	4814	99.9	922	5 AAB73838	Aab73838 Candida a
3	1679	34.8	859	2 AAW07872	Aaw07872 Aluminium
4	1666	34.6	858	2 AAW07873	Aaw07873 Aluminium
5	1099	22.8	663	6 ABJ25905	Abj25905 Aspergill
6	1099	22.8	663	6 ABJ25905	Abj25905 Aspergill
7	865	17.9	969	7 ADK62466	Adk62466 Disease t
8	821	17.0	228	5 ABP05970	Abp05970 Human ORF
9	345.5	7.2	154	5 ABP32160	Abp32160 Human gly
10	319.5	6.6	107	5 ABP06753	Abp06753 Human gly
11	251	5.2	326	9 AEM92441	Aem92441 M. xanthu
12	234.5	4.9	1198	4 ABB60371	Abb60371 Drosophil
13	210.5	4.4	354	9 AEB40726	Aeb40726 L. pneumo
14	210.5	4.4	357	9 AEB37409	Aeb37409 L. pneumo
15	186.5	3.9	1072	5 AEB54963	Aeb54963 Lactococc
16	182.5	3.8	373	4 AAC89814	Aag89814 C glutami
17	179.5	3.7	373	9 AED46971	Aed46971 Membrane
18	179.5	3.7	1172	5 ABG93308	Abg93308 C. albica
19	179	3.7	1199	4 ABB58274	Abb58274 Drosophil
20	175.5	3.6	1246	4 ABB61547	Abb61547 Drosophil

21	173.5	3.6	391	8 ADN47899	Adn47899 Thermococ
22	173	3.6	803	9 AEB53775	Aeb53775 Drosophil
23	170	3.5	1436	3 AAB18199	Aab18199 Plasmodiu
24	169	3.5	1468	4 ABB65329	Abb65329 Drosophil
25	168.5	3.5	816	6 AAE30116	Aae30116 Fruit fly
26	167.5	3.5	1458	10 AEF13870	Aef13870 Mouse NFA
27	166.5	3.5	1046	10 AEF38379	Aef38379 Dictyoste
28	166.5	3.5	1046	10 AEF38399	Aef38399 Dictyoste
29	166.5	3.5	2261	6 ABJ18914	Abj18914 Pathogen
30	166.5	3.5	2271	6 ABW72734	Abw72734 Staphyloc
31	166.5	3.5	2283	5 ABP56876	Abp56876 Staphyloc
32	166	3.4	688	5 ABP74039	Abp74039 Candida a
33	165.5	3.4	1977	7 ADE15647	Ade15647 Human str
34	165	3.4	2112	4 ABB60403	Abb60403 Drosophil
35	165	3.4	4498	4 ABB58595	Abb58595 Drosophil
36	164.5	3.4	2271	6 ABU16000	Abu16000 Protein e
37	161	3.3	366	4 AAG81142	Aag81142 Mycobacte
38	161	3.3	471	7 AD874419	Ad874419 Mycobacte
39	159.5	3.3	1169	6 ABU43255	Abu43255 Protein e
40	159.5	3.3	1177	5 ABP40261	Abp40261 Staphyloc
41	159.5	3.3	1177	8 ADS06718	Ads06718 Staphyloc
42	159.5	3.3	1253	3 AAB19772	Aab19772 Human den
43	159.5	3.3	1253	5 ABP51785	Abp51785 Human den
44	159.5	3.3	1253	9 ADY28712	Ady28712 Human pho
45	159.5	3.3	2150	5 AAO22566	Aao22566 Wooden le

ALIGNMENTS

RESULT 1

AAB36515

ID AAB36515 standard; protein; 922 AA.

XX

AC AAB36515;

XX

DT 06-MAR-2001 (first entry)

XX

DE Candida albicans CaALR1 protein sequence.

XX

EW Candida albicans; yeast pathogen; identification; fungal; antifungal;

KW CaKRE5; CaALR1; CaCDC24; diagnosis; fungicide; fungal infection.

XX

OS Candida albicans.

XX

FW WO200068420-A2.

XX

PD 16-NOV-2000.

XX

PP 05-MAY-2000; 2000WO-CA000533.

XX

PR 05-MAY-1999; 99US-0132878P.

XX

PA (MYCO-) MYCOTA BIOSCIENCES INC.

XX

PI Roemer T, Bussey H, Davison J;

XX

DR WPI; 2000-687652/67.

XX

DR N-PSDB; AAC87954.

XX

FT New DNA encoding essential proteins of Candida albicans, useful for

XX

PT diagnosing fungal infections and to screen for clinical or agricultural

XX

PS antifungal agents.

XX

PS Claim 14; Fig 2A; 76pp; English.

XX

CC The present sequence represents the Candida albicans CaALR1 protein. The

XX

CC present invention describes the fungus-specific genes CaKRE5, CaALR1 and

XX

CC CaCDC24 isolated in the yeast pathogen C. albicans. The genes have

XX

CC antifungal and fungicide activity. The genes in C. albicans and are

Db	781	NLLAYEKISRSHSNYLAQLQVESPNKNKITKTFSPKITLGTMLUPLNVLVTCLFGMNR	841
Qy	841	VPBGSGTNLGMWFGVIGVLPIIIGISPFIAQWLKKLNNISIQQNGNRPIFNHSSRSR	900
Db	841	VPBGSGTNLGMWFGVIGVLPIIIGISPFIAQWLKKLNNISIQQNGNRPIFNHSSRSR	900
Qy	901	RSLGKHKHGKNSIISFPNKYE	922
Db	901	RSLGKHKHGKNSIISFPNKYE	922
RESULT 2			
ID	ABF73838		
ID	ABF73838	standard; protein; 922 AA.	
XX	ABF73838;		
DT	30-JAN-2003	(first entry)	
XX			
DE		Candida albicans essential protein SEQ ID NO 7675.	
XX			
KW	Fungal; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;		
KW	signal transduction; DNA replication; cell division; growth;		
KW	proliferation; Candida albicans; fungicide; antifungal.		
OS	Candida albicans.		
XX			
PN	WO200253728-A2.		
PD	11-JUL-2002.		
XX			
PF	26-DEC-2001; 2001WO-US049486.		
PR	29-DEC-2000; 2000US-0259128P.		
PR	20-FEB-2001; 2001US-00792024.		
PR	22-AUG-2001; 2001US-0314050P.		
XX			
PA	(ELIT-) ELITRA PHARM INC.		
XX			
PI	Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;		
DR	WPI; 2002-566694/60.		
DR	N-PSDB; ABZ32388.		
XX			
PT	Constructing strains for identifying gene products as effective targets		
PT	for therapeutic intervention, by inactivating in the strain one allele of		
PT	a gene and placing other allele of the gene under conditional expression.		
XX			
XX	Claim 44; SEQ ID NO 7675; 167bp + Sequence Listing; English.		
PS			
CC	The invention relates to constructing (M1) a strain of diploid fungal		
CC	cells in which both alleles of a gene are modified, comprising modifying		
CC	one allele by insertion or replacement by a cassette having an		
CC	expressible selectable marker and modifying other allele by		
CC	recombination, of a promoter replacement fragment with a heterologous		
CC	promoter, so that expression of the second allele is regulated by the		
CC	promoter. (M1) is useful for constructing a strain of diploid fungal		
CC	cells in which both alleles of a gene are modified. The diploid fungal		
CC	cells having both alleles modified are useful for identifying a gene that		
CC	is essential to the survival or growth of a fungus, a gene that		
CC	contributes to the virulence and/or pathogenicity of a fungus, a gene		
CC	that contributes to the resistance of a diploid fungus to an antifungal		
CC	agent, an antifungal agent that inhibits the growth of a diploid fungus		
CC	and for identifying a therapeutic agent for treatment of a mammalian		
CC	disease. (M1) is useful for identifying a compound which modulates the		
CC	activity of a gene product, preferably enzymatic activity, carbon		
CC	compound catabolism, biosynthetic, transporter, transcriptional.		
CC	translational, signal transduction, DNA replication and cell division		
CC	activity. The method is useful for identifying a compound having the		
CC	ability to inhibit growth or proliferation of C. albicans cells and for		

CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office  
XX  
XX  
5Q Sequence 922 AA;  
Query Match 99.9%; Score 4814; DB 5; Length 922;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 921; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 MSDSYQNSTNQIPRSDVLDHRNQITNDCAISDELELSELESEVVKSEKQ 60  
Db 1 MSDSYQNSTNQIPRSDVLDHRNQITNDCAISDELELSELESEVVKSEKQ 60  
Qy 61 QQHGEITSDNAKPLTRKSGSIKKSNLTKDKRITNPKSLSGDDTINSGHKNRYNMS 120  
Db 61 QQHGEITSDNAKPLTRKSGSIKKSNLTKDKRITNPKSLSGDDTINSGHKNRYNMS 120  
Qy 121 SLRKDFYKNDTNDSTNNHTLAIPPIPIPIITNANKSRKSKOLENPLPKKTYI 180  
Db 121 SLRKDFYKNDTNDSTNNHTLAIPPIPIPIITNANKSRKSKOLENPLPKKTYI 180  
Qy 181 GRNSNPFNDLVSPMTKKTNDSEDTNTSTTANRMLGIGATTLGVTGTTATATA 240  
Db 181 GRNSNPFNDLVSPMTKKTNDSEDTNTSTTANRMLGIGATTLGVTGTTATATA 240  
Qy 241 AAGRPSRSIDSADSHASRSQTEEDVCFPMVGDHVRVNGIDFIDEFIREESEA 300  
Db 241 AAGRPSRSIDSADSHASRSQTEEDVCFPMVGDHVRVNGIDFIDEFIREESEA 300  
Qy 301 YLQKMTAKNLRIDFQNLKNNNTSGASRPHYHHSNNKNNKNGDGGSSNAALKTY 360  
Db 301 YLQKMTAKNLRIDFQNLKNNNTSGASRPHYHHSNNKNNKNGDGGSSNAALKTY 360  
Qy 361 PKNLLKTLSPETHSSSEIEYELTKQPPYDDQLSLTSTSTSGSGSQVK 420  
Db 361 PKNLLKTLSPETHSSSEIEYELTKQPPYDDQLSLTSTSTSGSGSQVK 420  
Qy 421 FGARISDINGSLPDRPSLPHSESETHAPDIPSLVSPQSVRLPFGEEETWLD 480  
Db 421 FGARISDINGSLPDRPSLPHSESETHAPDIPSLVSPQSVRLPFGEEETWLD 480  
Qy 481 TCTDSMKLAKAFGIHPLTAEDIRMQETREKVELFKSYVFCFHTFEADKESDYLEP 540  
Db 481 TCTDSMKLAKAFGIHPLTAEDIRMQETREKVELFKSYVFCFHTFEADKESDYLEP 540  
Qy 541 INVYIVPHDGLTFHPSLPHANVRVRQLRDYVDSADMLCYALIDEITDGFAPVI 600  
Db 541 INVYIVPHDGLTFHPSLPHANVRVRQLRDYVDSADMLCYALIDEITDGFAPVI 600  
Qy 601 HGIEYADAEIVFTARDTDFSSNLQIGESRRKVTMLRLLSKADVIKFAKRCQEE 660  
Db 601 HGIEYADAEIVFTARDTDFSSNLQIGESRRKVTMLRLLSKADVIKFAKRCQEE 660  
Qy 661 ANSSGGYQRYNQV 720  
Db 661 ANSSGGYQRYNQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 720  
Qy 721 NPTGNNTNTNTTSGSPSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 780  
Db 721 NPTGNNTNTNTTSGSPSPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 780  
Qy 781 NLLAYEKIFSRSHSNYLAQLQVESFNSNNKITEMFSKILIGTMLVPLVGLFGNVR 840  
Db 781 NLLAYEKIFSRSHSNYLAQLQVESFNSNNKITEMFSKILIGTMLVPLVGLFGNVR 840  
Qy 841 VPGEGGTLNLFVGLVFIIFIGSFIFAQWLKLNNSIEQNGNRPFNHSSRSI 900  
Db 841 VPGEGGTLNLFVGLVFIIFIGSFIFAQWLKLNNSIEQNGNRPFNHSSRSI 900

Db 841 VPGEGGTLNLFVGLVFIIFIGSFIFAQWLKLNNSIEQNGNRPFNHSSRSI 900  
Qy 901 RSLGLKKGKNSIISPPNKYE 922  
Db 901 RSLGLKKGKNSIISPPNKYE 922  
RESULT 3  
AAW07872  
ID AAW07872 standard; protein: 859 AA.  
XX  
XX AAW07872;  
XX 23-FEB-1997 (first entry)  
DE Aluminium resistance gene ALR1.  
XX Aluminium resistance; ALR1; ALR2; ARH1; cation; tolerance;  
XX magnesium transport.  
XX Saccharomyces cerevisiae.  
XX W09634959-A1.  
XX 07-NOV-1996.  
XX 01-MAY-1996; 96WO-NZ000035.  
XX 01-MAY-1995; 95NZ-00272039.  
XX (AUCK-) AUCKLAND UNISERVICES LTD.  
XX Gardner RC, Macdiarmid CW, Hay RJM;  
XX WPI; 1996-506161/50.  
XX ARL1-homologues, aluminium resistance genes ALR1 and ALR2 - used in the  
treatment of cation deficiency, cation toxicity, esp. heart disease.  
XX Claim 6; Fig 5; 62pp; English.  
XX ALR1 and ALR2 are isolated by selecting vectors which confer high  
aluminium tolerance, or that complement knockout mutations in ALR1, ALR2  
and/or ARH1. The genes have homology with bacterial genes responsible for  
divalent ion uptake. The ALR1 and ALR2 are magnesium transport genes. The  
isolated transport genes can be used in the treatment of any plant,  
animal or microorganism disease which results from a cation deficiency,  
by producing an accumulation of cations in a plant deficient to those  
cations, or in plants consumed by animals deficient in those cations. The  
genes may also be used in the treatment of diseases resulting from cation  
toxicity, esp. manganese toxicity and esp. for the treatment of heart  
disease. Note: The ARL1 and ARL2 gene sequences are claimed and stated as  
given in Figure 6 and Figure 7 of the specification, however these  
figures are missing from the specification  
XX  
XX Sequence 859 AA;  
Query Match 34.88; Score 1679; DB 2; Length 859;  
Best Local Similarity 41.5%; Pred. No. 6.2e-119;  
Matches 405; Conservative 137; Mismatches 258; Indels 176; Gaps 31;  
Qy 1 MSDSYQNSTNQIPRSDVLDHRNQITNDCAISDELELSELESEVVKSEKQ 57  
Db 1 MSSSSS--SSSESPYLSRS-----NSLANTWVSMKTEDHTGLYDHRQHPDLSLVRH 49  
Qy 58 EQQOQHGEITSDNAKPLTRKSGSIKKKS-----NLTKDRITNPKSLSGDDTINSGH 112  
Db 50 QPPTLKKEKSTKPSIKPEKQKATYNSHYDVGVSFSGRMDFDEQGMDETVA-AHH 108  
Qy 113 KNR-----NTNMSSSI-----RKDFYKNDTNDSTNNHT---HLAIPPIPIPTPII 156



PR	31-AUG-2001; 200IUS-0316362P.
PA	(ELIT-) ELITRA PHARM INC.
PI	Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
PT	WPI; 2003-093124/08.
DR	
XX	
PT	New purified or isolated nucleic acids of essential genes of Aspergillus
PT	fumigatus, useful for treating or preventing infections by A. fumigatus,
PT	or for treating a non-infectious disease in a subject e.g. cancer.
XX	
PT	Disclosure; Page; 175pp; English.
XX	
CC	The invention relates to novel purified or isolated nucleic acids of
CC	essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC	the invention are used to treat or prevent infections by a pathogenic
CC	organism such as A. fumigatus, to treat a non-infectious disease in a
CC	subject (e.g. cancer), to prevent or contain contamination of an object
CC	by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC	biofilm comprising A. fumigatus. The polynucleotides are useful for
CC	expressing recombinant protein for characterisation, screening or
CC	therapeutic use, as markers for host tissues in which the pathogenic
CC	organisms invade or reside, for comparing with the DNA sequence of A.
CC	fungus to identify duplicated genes or paralogues having the same or
CC	similar biochemical activity and/or function, for comparing with DNA
CC	sequences of other related or distant pathogenic organisms to identify
CC	potential orthologous essential or virulence genes, for selecting and
CC	making oligomers for attachment to a nucleic acid array for examination
CC	of expression patterns, for raising anti-protein antibodies, as an
CC	antigen to raise anti-DNA antibodies or to elicit another immune
CC	response, and for identifying polynucleotides encoding the other protein
CC	with which binding occurs or to identify inhibitors of the binding
CC	interaction. The polypeptides may be used to raise antibodies or to
CC	elicit immune response, as a reagent in assays designed to quantitatively
CC	determine levels of the protein in biological fluids, as a marker for
CC	host tissues in which pathogenic organism invade or reside, and to
CC	isolate correlative receptors or ligands in the case of virulence
CC	factors. This sequence represents a protein of one of the essential genes
CC	of Aspergillus fumigatus of the invention
XX	
SQ	Sequence 663 AA:
	Query Match            22.8%; Score 1099; DB 6; Length 663;
	Best Local Similarity 32.7%; Pred. No. 9.7e-75;
	Matches 292; Conservative 98; Mismatches 241; Indels 262; Gaps 22;
QY	4 SESYQNSTNQIPRSEVLDHRRNQYNQCAISDESELELKSLSBSEVVKSEQQQH 63
DB	7 SEXMYCLTRFTVPVE---LDHHRQL-----DSPPREATADI-SLRGNQAQHY 54
QY	64 HQEITSDNAKLTKSGSISKKKSNITDKDIRNPKMSLGSGDDTINSCHKRNYYNSSLR 123
DB	55 HQETPQ-----RPDLSIQDALREAGSLSRDPEAQILD 87
QY	124 KDFYLKD-----NYDNSTNNH-----THLAIPPIPIPTIITNANKSPRK--SQ 167
DB	88 DNRSGKDINTLGRFSPVDNGNVRGRTWSKHQEL-----ANNSRESGSPA 134
QY	168 LEXMLPLIKKTIGRRNNNNPNENDLVSPMTCKNTDSEDITWTSTTTANMKLGTCATTG 227
DB	135 RSSSPR-----NSEVAFAFPRRERANTLE-----SHAAPDLAAILQRTVSG 176
QY	228 VGTGT-TATATAAAAGRPSRSIDSADSHASRSQTEDVCFPVMGDHVRNGIDP 286
DB	177 ---GTHPRPPTFNMAARPOPGDIQLEN-----DESCVTTFEQGRIPVIDY 222
QY	287 DEIDEFIREEREAYLOQMXTAKNILLRIDEFQNLKNNTTSGSARHP---YHHHSNNKK 343
DB	223 BELKEEFA-----LSRMKPSTSR-----KQSLSSQSGRGPFYDLRPLGRK 266

QY	344	NNGGDGGSGSMAALKYTPKNILKTLSPRFTFTHNSSSSEIYELKTKQPPKYVDOL	403
Db	267	DVEGKSSSA-----DRSSDLMADLUKTAUKYANVVDK	304
QY	404	LTSSTSTSGSGSQVKGGAISDINGGSLPDRFLPHSBSBETHADIPSLVSPQ	463
Db	305	-----IWEKQNEBFTFQFFSSSSQSTVHAAELGDLVPGD	342
QY	464	SVRDLPRNGE--TWMLDCTCPTDSEKMKLAKGCIHPLTARDIRMOETREKVELPKSY	521
Db	343	TRFDQLGPEGVWDLVNLNTEEEVAALSRFSINPLTIEDILUTQAREKVELFKQY	402
QY	522	FVCFHTF-EADKESDYLEPINVYVFDHGILTFHFSPISHPNVRRVRLQDYDVS	580
Db	403	FVCFRTFQLDKTDERPMFVNFYVVRDGLSFSTENPHAAVNRKIGRLDYLS	462
QY	581	ADMCLYALDELTGCPAVTHGIEVEADAEVFTARDTDFSSMLORIGESRKYVTLM	640
Db	463	SDWICYAMIDDDVDSFGVIREIESEAEIDLVIARVDDPFSPLRIGGURKKVMSLM	522
QY	641	RLLSGKADVIKMFARQCEANSSSGYYQRYNLQOQQOQAPPPNPITSPINSTLNL	700
Db	523	RLLSGKADVIRGFSRCNE-----	541
QY	701	NSLGTSGGCVGCGINFGPNTGNNTNTNTTSGSPPOOQOQHITNKSPFPDAP	760
Db	542	-----QYSVT-----P	547
QY	761	RADIALYLGDIQDHIITWQNLAYEKIPFSRSHSNLYLAQLQVESFNSNKNKITEMFSKITL	820
Db	548	RGDIGLYGDIQDHWVTMSNLAFHFKMLRSRHTNYLAQLVNLVLGNHNVKLSKVTL	607
QY	821	ICTMLVPLNVLVGLFGNVRVPGCGTNLWGFVGIVGLYFIIGSIFPAQW	873
Db	608	IATMLVPMNLICGLFGMNVVRVPGCGIGLWGFVGIVGVIAAVILUSGLAARY	660
RESULT 6			
ABJ76505	ABJ76505 standard; protein; 663 AA.		
XX	DT	ABJ76505	
XX	AC	ABJ76505	
XX	XX		
XX	DE	16-APR-2003 (first entry)	
XX	DE	Aspergillus fumigatus essential gene protein #1163.	
XX	KW	Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;	
XX	KW	cancer; contamination; biofilm; antibody; immune response.	
XX	OS	Aspergillus fumigatus.	
XX	XX	WO200286090-A2.	
XX	PN		
XX	XX	31-OCT-2002.	
XX	PD		
XX	PP	23-APR-2002; 2002WO-US013142.	
XX	XX		
PR	PR	23-APR-2001; 2001US-0285697P.	
PR	PR	27-APR-2001; 2001US-0287066P.	
PR	PR	05-JUN-2001; 2001US-0295890P.	
PR	PR	09-JUL-2001; 2001US-0301899P.	
PR	PR	31-AUG-2001; 2001US-0316362P.	
XX	XX	(ELIT-) ELITRA PHARM INC.	
XX	FA		
XX	XX	Jiang B, Tishkoff D, Zamudio C, Broshkin AM, Hu W, Lemieux SM;	
XX	PI		
XX	XX	WPI; 2003-093124/08.	
DR	XX		

XX		New purified or isolated nucleic acids of essential genes of Aspergillus	
PT		fumigatus, useful for treating or preventing infections by A. fumigatus,	
PT		or for treating a non-infectious disease in a subject e.g. cancer.	
XX		Disclosure; Page: 175pp; English.	
PS			
XX		The invention relates to novel purified or isolated nucleic acids of	
XX		essential genes of Aspergillus fumigatus. The isolated nucleic acids of	
CC		the invention are used to treat or prevent infections by a pathogenic	
CC		organism such as A. fumigatus, to treat a non-infectious disease in a	
CC		subject (e.g. cancer), to prevent or contain contamination of an object	
CC		by A. fumigatus, or to prevent or inhibit formation on a surface of a	
CC		biofilm comprising A. fumigatus. The polynucleotides are useful for	
CC		expressing recombinant protein for characterisation, screening or	
CC		therapeutic use, as markers for host tissues in which the pathogenic	
CC		organisms invade or reside, for comparing with the DNA sequence of A.	
CC		fumigatus to identify duplicated genes or paralogues having the same or	
CC		similar biochemical activity and/or function, for comparing with DNA	
CC		sequences of other related or distant pathogenic organisms to identify	
CC		potential orthologous essential or virulence genes, for selecting and	
CC		making oligomers for attachment to a nucleic acid array for examination	
CC		of expression patterns, for raising anti-protein antibodies, as an	
CC		antigen to raise anti-DNA antibodies or to elicit another immune	
CC		response, and for identifying polynucleotides encoding the other protein	
CC		with which binding occurs or to identify inhibitors of the binding	
CC		interaction. The polypeptides may be used to raise antibodies or to	
CC		elicit immune response, as a reagent in assays designed to quantitatively	
CC		determine levels of the protein in biological fluids, as a marker for	
CC		host tissues in which pathogenic organism invade or reside, and to	
CC		isolate correlative receptors or ligands in the case of virulence	
CC		factors. This sequence represents a protein of one of the essential genes	
CC		of Aspergillus fumigatus of the invention	
XX		Sequence 663 AA;	
SQ			
Query Match 22.8%; Score 1099; DB 6; Length 663;			
Best Local Similarity 32.7%; Pred. No. 9.7e-75;			
Matches 292; Conservative 98; Mismatches 241; Indels 262; Gaps 22;			
QY	4	SESYQNSTNOPIPRSDVLDHNNQITNDCAISDESELEKSESEVVKSKQOQH	63
Db	7	SEKYNCLTREFTPVE-----LDDHFPQ-----DPPRIATADI-SLSQNTAQHAY	54
QY	64	HOEITSDNAKPLTRKSGSSIKKSNLTKDRITNPMISLGGDDTINSGHKNRNNYNSLR	123
Db	55	HOETPQ-----RPDLLSIQDALREAGSLSRDPQAILD	87
QY	124	KDFYIKD-----NTDDNSTNNH-----THLAIPPIPIPTIITNANKSRK--SQ	167
Db	88	DRSGKQINTLGRFSDPNGNVRHGRWTGRTHQEL-----ANMSRESFSA	134
QY	168	LENLPPLIKKTKTIGRNNNFENDLVSPMTKMTNDSEDITNTSTTANHKLGIGATTLG	227
Db	135	RSSSP-----NSVEAFADPRRERANTLE-----SHAAPDLAILQRTVSG	176
QY	228	VGTGT-TATATATAAGRPSRSSIDSEADSHASRSQETEDVCFPMVGDHVRVNGIDF	286
Db	177	---GTHPRPTFSNAIRPQPGDIOLEPN-----DESCVPTTQGPRIPIVDY	222
QY	287	DEIDFIREEREAYLOKQIAKNILRIDFONLSKNTTSGASRHP---YHHHNNKK	343
Db	223	ELEBEFA-----LSRQMKPSTSR-----KQSLSSQSGRPVYDLRPLGRKS	266
QY	344	NNGGDGGSGSMAALKYTPKNILKTLSPRFTFTHNSSSSEIYELKTKQPPKYVDOL	403
Db	267	DVEGKSSSA-----DRSSDLMADLUKTAUKYANVVDK	304
QY	404	LTSSTSTSGSGSQVKGGAISDINGGSLPDRFLPHSBSBETHADIPSLVSPQ	463

Db 305 -----IIVEKIQNEPTEFRGFFSESQSTVHAAELGDLVLPD 342

Qy 464 SVRLDFRNGEE--TWLDCCTPTDSEKMLAKAFGIHPLTAEDIRMOETREKVELFKSY 521

Db 343 TFDLPLQGFEGGVWMLDLNPTSEVAALSRAFSIHPLTIEDIITQAREKVELFKQY 402

Qy 522 FVCFTPT-EADKESDYLEPINVIYVFDGILTFHSPISHPNVRRVRLQRYDVVS 580

Db 403 FVCFTPTFQDLKDFRPFEPVNFYVVFRCVLSFSEFENPHAAANVRKRGKLDYVLS 462

Qy 581 ADWLCVALIDEITDGFAPVTHGIEYEDADEVFTARDTFSSMLQRICESRRKVTLM 640

Db 463 SDWICYAMIDIVDFGVREIESEATEDVFIARVDVFESFLPRIGLURKVMSLM 522

Qy 641 RLSSKADVIMFKAKROEBANSSGGYQRYNLOQQOQQOQPPPPPIITSPINSTLNL 700

Db 523 RLGGKADVIRGFKRCNE-----541

Qy 701 NSLGTSTGGGVGGVGNFGNPTGNNTNTNTTNGSPSPQOQOQHGHTNKSFPIDARP 760

Db 542 -----QYSVT-----P 547

Qy 761 RADIALYLGDIQHIITMQLLAYEKIFSRSHSNLYLAQLOVESFNNSKITEMFSKITL 820

Db 548 RGDIGLYGDIQDHYVVTWMSNLAFHFKMLSRSHNTYLAQLNVTNLVLGNHVNKLSKVTL 607

Qy 821 IGTMLVPLNLVTGLFGNNRVPGEGTNLQWFFGIVGLVFIILIGSEIFAQW 873

Db 608 IATMLVPMNLICGLFGNNRVPGEGTGLQWFFGIVGVIAAVLVLGTAARY 660

RESULT 7

ADK62466

ID ADK62466 standard; protein; 969 AA.

XX

AC ADK62466;

XX

DT 06-MAY-2004 (first entry)

XX

DE Disease treating protein complex-derived protein #359.

XX

KW protein complex; drug target; diagnosis.

XX

OS Unidentified.

XX

PN EP1338608-A2.

XX

PD 27-AUG-2003.

XX

PP 20-DEC-2002; 2002EP-00102902.

XX

PR 20-DEC-2001; 2001EP-00130253.

XX

PA (CELL-) CELLZOME AG.

XX

PI Bauer A, Gavin A, Superti-Furga G, Kuester B, Schultz J;

PI Marzoch M, Grandi P, Krause R, Kruse U, Merino A, Bauch A;

PI Michon A, Leutwein C, Rick J;

XX

DR WPI; 2003-638460/61.

XX

DR N-PSDB; ADK62467.

XX

PT New proteins and protein complexes from eukaryotes, useful as targets in

PT drug screening, or in diagnosing or screening for the presence of a

PT disease or disorder, or a predisposition for developing a disease or

PT disorder in a subject.

XX

PS Disclosure; SEQ ID NO 717; 13pp; English.

XX

XX The invention relates to novel protein complexes comprising a first and a

CC second protein, or its derivative, fragment, homologue or variant. The

CC proteins are selected from given protein complexes, which are not defined

CC in the specification. The variants are encoded by nucleic acids that

CC hybridize to the nucleic acids encoding the proteins under low stringency

CC conditions. The protein complexes are useful as targets for an active

CC agent of a pharmaceutical. These protein complexes are particularly

CC useful as drugs targets for the treatment or preventing of a disease or

CC disorder. The complexes and methods above are useful in diagnosing or

CC screening for the presence of a disease or disorder or a predisposition

CC for developing a disease or disorder in a subject. These are also useful

CC in screening for a drug for treatment or prevention of a disease or

CC disorder. The molecule that modulates the amount, activity or protein

CC components of the complex is useful for the manufacture of a medicament

CC for the treatment or prevention of a disease or disorder. This sequence

CC corresponds to a protein of the invention. (Note: the sequence data for

CC this patent did not form part of the printed specification but was

CC obtained from the EPO in electronic format).

XX

Sequence 969 AA;

Query Match 17.9%; Score 865; DB 7; Length 969;

Best Local Similarity 24.2%; Pred. No. 1.3e-56;

Matches 271; Conservative 143; Mismatches 247; Indels 460; Gaps 29;

Qy 11 STTNOPISRDEVLDDHRNQITNDCAISDSEDELELKLSESEVVKSE-----58

Db 44 SDSRRPTQLLDHNLQHNHGGITDPTQI-DSWGLHSHSDSTNDLIKSEDSPLKGAFIGHR 102

Qy 59 ----KQOQHQEITSD-NAKPLTRKSGSSIKKKNLTKDKRITNPMG-LSGSDDTINSGH 112

Db 103 PMSQREGPQSVSTVQPPQIPMKFSTPSYKKPAGLRPSQNRSLVSLDSPSE--LESWL 160

Qy 113 KNRNNTSSLRKDFYLDN--TDDNSTNNHTLAIPPIPIPTPIITNA-----NKRNRK 165

Db 161 KRR-----KSVHKSFDENSPTDRQSNANDVDVVD-ALMNVNNASTCYVNDKRRK 215

Qy 166 SQLENPLPKKTKTGRNNPNFENDLVSPMTOMKTNDSEDITNTSTANHMKLGIGATT 225

Db 216 -----KKR--GSDSSN-KN--SKSTSSDSNDEDEYNS-----244

Qy 226 LGVGTGTATATATAAGRPGRSSIDSEADSHASRSSOETEDVCPM--VGDHVRWG 283

Db 245 -----RPSS-----LSSNSSLDDVCLVLDDEGSEVPKAW 275

Qy 284 IDPDEIDPIREEREAYLQKQIAKNIILRIDEFONLSKNNTTSGASRHPYHHNNKK 343

Db 276 PDCTVLEEPSKEETER--LRSQAI-----QDAEAPHQYDEDEE 312

Qy 344 NNGDGGGSSMAALATPKNLKLTLSRPEFTHESSSSSEIVELTQOPPVYDQLS 403

Db 313 D-----GTSNEDGILFSPKPIVNIIDVFNLRVNYTENLKNGLRPRFRIAPHLIQPM 367

Qy 404 LTSSTSTSGSGSGGVKFGGARISDGINGGSL-----PD--RPSLFHSES 446

Db 368 VL-----GSNSTKDSKRSIQSLQDNLVGNRTQYPPHISNNPHFRTYFRVDL 418

Qy 447 EETTHAPDIPSLVSPQSVRDLP-----ADKES-----469

Db 419 DSTVHSITISGLLQPGQFDLIVASISQDNSAGHKHTHPSPTPGIKAEVTSLOQLT 478

Qy 470 -----RNGEET--WWLDCCTPTDSEKMLAKAFGIHPLTAEDIRMOETREKVE 515

Db 479 AKNPSTLSSNSVANTIEDVPPFLVDVSNPTEEMKILSKAFGIHPLTTEDIFLGEVREKVE 538

Qy 516 LFKSYTVVCPHPTF-----ADKES-----534

Db 539 LFRDYTLICFRSFDIVAEKHVRRRREKQESATLDHESISRRKSKQAYGATMNSNANN 598

Qy 535 -----534

Db 599 NSTSNASRKLWFLSLARRRRSSANRTTNTSSSYKRVKSEKKKQENEFKFKSGDRH 658  
QY 535 ---EDYLEPNIYIVVPHDGLTFPHFSPISHPANVRRVRQLRDYVDVSDMLCYALIDE 591  
Db 659 KPEGELEPLNVIIVFTGVLTFHPTFHPINVRRAELKDLVNTVSDIAIALDD 718  
QY 592 ITGDPAPVINGIEYADAEIDAVFTARDTFSS----- 624  
Db 719 ITDAFAPMBLIEDEVEIEDAILKMHQSDSDSDSDSDSDSDSDSDSDSDSDSDSD 778  
QY 625 -----MLQIGESRRKVTMLRLLSKADVI 650  
Db 779 TSYSSAKSSVSRSMTSEASFNALIGWRKGMRLRIGCKRVMSILRLSGADVI 838  
QY 651 KPAKRCQBEANSSGGYVQYQVNLQOQQOQAPPPPNPIITSPINSLNLSLGTSGG 710  
Db 839 KGAKR-----YNEQW----- 849  
QY 711 VGVGGINFGPNPTGNTNTNTTGTSPSPQOQOQHGTWKSPFPDARPRADIALYGD 770  
Db 850 -----EASQSEIARYLGD 863  
QY 771 IQDHIITMQLLAYEKIFSRSHSNLYLAQLQVESFNNNKITEMFSKITLIGTMLVPLN 830  
Db 864 IQDHIITMQLLAYEKIFSRSHSNLYLAQLQVESFNNNKITEMFSKITLIGTMLVPLN 923  
QY 831 VTGLFCMNVVRPCEGNTLQWPGVIGVLFPIIGSFPAQ 871  
Db 924 ITGLMGRNVIVPQYRDLSTWFIIGVLFMCHLACSAVNTK 964

RESULT 8

ABP05970  
ID ABP05970 standard; protein; 228 AA.  
XX  
AC ABP05970;  
XX  
XX 24-JUN-2002 (first entry)  
XX  
DE Human ORFX protein sequence SEQ ID NO:11922.  
XX  
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
KW hypertension; hypothyroidism; cholesterol ester storage disease;  
KW immune deficiency; immune disorder; infectious disease;  
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
KW myasthenia gravis.  
XX  
OS Homo sapiens.  
XX  
PN WO200192523-A2.  
XX  
XX 06-DEC-2001.  
XX  
XX 29-MAY-2001; 2001WO-US010836.  
XX  
PR 30-MAY-2000; 2000US-0206132P.  
PR 29-AUG-2000; 2000US-0228716P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Shinketsu RA, Leach MD;  
XX  
DR WPI; 2002-106308/14.  
DR N-PSDB; ABN21722.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,

hyperproliferative disorders and autoimmune disorders.  
XX  
XX Disclosure; SEQ ID NO 11922; 1037pp; English.  
XX  
CC The present invention describes substantially purified human proteins  
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN15762 to ABN27252 encode the human ORFX  
CC proteins given in ABP0010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 228 AA;  
Query Match 17.0%; Score 821; DB 5; Length 228;  
Best Local Similarity 59.6%; Pred. No. 3.8e-54;  
Matches 170; Conservative 23; Mismatches 32; Indels 60; Gaps 5;  
QY 510 TREKVELFKSYFVFCFTPEADKESDYLEPINVIYVPHDGLTFPHFSPISHPANVRR 569  
Db 1 TREKVELFKSYFVFCFTPEADKESDYLEPINVIYVPHDGLTFPHFSPISHPANVRR 60  
QY 570 VRQLRDYVDVSDMLCYALIDEITDGFAPVINGIEYADAEIDAVFTARDTFSSMLQRI 629  
Db 61 VRQLRDYVDVSDMLCYALIDEITDGFAPVINGIEYADAEIDAVFTARDTFSSMLQRI 120  
QY 630 GSRKRVMTMLRLLSKADVIKMAKRCQBEANSSGGYVQYQYQVNLQOQQOQAPPPPP 686  
Db 121 GSRKRVMTMLRLLSKADVIKMAKRCQBEANSSGGYVQYQYQVNLQOQQOQAPPPPP 173  
QY 687 NPITSPINSLNLSLGTSGTGGVGVGINFGPNPTGNTNTNTTGTSPSPQOQOQH 746  
Db 174 -----NVNQ-----NNSN-----NQI 184  
QY 747 GITNKSFPIDARPRADIALYGDIDQHIITMQLLAYEKIFSR 791  
Db 185 SILSN-SYMOSTQPRGDIALYGDIDQHIITMQLLAYEKIFSR 228  
RESULT 9  
ABP32160  
ID ABP32160 standard; protein; 154 AA.  
XX  
XX ABP32160;  
XX  
XX 08-JUL-2002 (first entry)  
XX  
XX Human glycoprotein-like ORF1133 protein, SEQ ID NO:2266.  
XX  
KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
KW immune modulation; hematopoiesis regulation; tissue growth;  
KW angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;





psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

5Q	Sequence 107 AA;	Query Match	6.6%;	Score 319.5;	DB 5;	Length 107;
		Best Local Similarity	62.9%;	Pred. No. 2.8e-16;		
		Matches	56;	Conservative	21;	Mismatches 11;
						Indels 1;
						Gaps 1;

	Matches	56;	Conservative	21;	Mismatches	11;	Indels	1;	Gaps	1;
QY	790	SRSHSYLAQLQVESFNSNNKITEMFSKITLIGTILGMLPLVPLVTGLFGNNVRVPEGGTVL	849							
DB	1	SRSHSYLAQLQVESFNSNNKITEMFGKVTIGTILGMLPLVPLVTGLFGNNVRVPOE--NSSI	59							

```
Qy      850 GWFFGIVGVLIIFIIGSFIFAQWNLKLN 878  
        |:|||:||| : : |:|||:::  
Db      60 AWWFGILCVLLLLAVLGWFLASYTWIKRID 88
```

RESULT 11  
AEM92441  
ID AEM92441 standard; protein; 326 AA.

ABW92441;	AC
XX	
DT	02-JUN-2005 (first entry)

DE M. xanthus protein sequence, seq id 11640.  
XX  
KW Transgenic plant; DNA replication; gene regulation; gene expression.

OS	<i>Myxococcus xanthus</i> .
XX	
PN	US6833447-B1.

PD 21-DEC-2004.  
XX  
PF 10-JUL-2001; 2001US-00902540.

PR 10-JUL-2000; 2000US-0217883P.  
XX  
PA (MONS ) MONSANTO TECHNOLOGY LLC.

PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;  
XX  
DR WPI; 2005-028716/03.

PT New substantially purified *Myxococcus xanthus* nucleic acid molecule  
PT encoding a nitrite reductase, useful for determining gene expression,  
PT identifying mutations in a gene of interest, and for constructing

XX  
PS Example 2; SEQ ID NO 11640; 25pp; English.  
XX

CC encoding a nitrite reductase of SEQ ID NO. 11976. Further disclosed is a  
CC recombinant DNA construct for expression of a nitrite reductase gene in a  
CC plant cell, and a plant cell comprising the recombinant DNA construct.

Qy	449	TIHAPDIPSLVSPQSVDRLFRNGEETWMLDCTCPDTSEMKHLKAPGI-----	497
Dd	612	-----QAASSGNGN-----PNAIGMSSGVNGSGGAGNGAGGN	648
Qy	498	----HPLTAEDIRMQETREKVE---LFKSYYVFCFHTPFADKESEDYLSPIINVIVVFHD	550
Dd	649	SGPCNPWGTSATPGHGGEVIDPKHFLBELCVC-----GDKVSG-----	691
Qy	551	GILTFHFPSISHPANVRBRVQLDDVDVSADMLCYALDEITOGFAPVIH-----	603
Dd	692	GILLTCE-----SKGFFRRVQNKKVTVCY--AERSCH--IDKTQRKRCPYCRPOKCLEVM	744
Qy	604	EYEADAIEDAVFTARDTPDSMLQR-----IGESRXVMTLMRLLSKADVTKMFARCO	658
Dd	745	KLEA-VRADRMGRGN-KFGPMYKRDARKLQVMRQRLAQALRNMGPDIK-----	795
Qy	659	EEANSSSGYGYORYNIQQOQC-----QAPPPPNPIITSPINLTNLNSLGSTG	708
Dd	796	-PTPIFGVQAYNNIKNQEIQIPQVSSLTQSDSPSPI-----NALQVNASTG	847
Qy	709	-----GGVGCGGIFNGPNTGNNTNTNTTWS	736
Dd	848	GVIATPMNAGTCGGGGGLN-GPSSVCGNGSSNGSN	885
 RESULT 13 AEB40726			
ID	AEB40726	standard; protein; 354 AA.	
AC	AC	AEB40726;	
XX	XX		
DT	08-SEP-2005	(first entry)	
DE	L. pneumophila	protein SEQ ID NO 5058.	
DE	detection;	infection; Antibacterial; Vaccine.	
KW	Legionella	pneumophila.	
OS	XX	WO2005049642-A2.	
FN	XX		
PD	02-JUN-2005.		
PF	23-SEP-2004;	2004WO-IB003578.	
PR	21-NOV-2003;	2003FR-00013687.	
PA	(INSP )	INST PASTEUR.	
PA	(INRM )	INSERM INST NAT SANTE & RECH MEDICALE.	
PA	(UYLY-) UNIV LYON 1	BERNARD CLAUDE.	
PA	(CNRS )	CNRS CENT NAT RECH SCI.	
PI	Buchrieser C,	Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;	
PI	Rueniock C,	Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch P;	
PI	Jarraud S;		
DR	WPI:	2005-388305/40.	
XX	New genome of Legionella pneumophila Paris strain and derived		
XX	polypeptides, useful for detection or identification of the strain and		
XX	for treatment and prevention of infections.		
XX	Claim 3; SEQ ID NO 5058; 660pp; English.		
XX	The invention relates to an isolated or purified nucleotide sequences (I)		
CC	from Legionella pneumophila Paris strain. (II), and their related		
CC	sequences or fragments, are useful as primers and probes for detection		
CC	and amplification, including differentiation between the Paris and		
CC	Philadelphia strains of Legionella pneumophila and to prepare recombinant		
CC	(hybrid) polypeptides (II). (II) are also useful for preparation of		



DE Lactococcus lactis protein yqfG.  
XX  
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX  
OS Lactococcus lactis; IL1403.  
XX  
PN FR2807446-A1.  
XX  
PD 12-OCT-2001.  
XX  
PF 11-APR-2000; 2000PR-00004630.  
XX  
PR 11-APR-2000; 2000PR-00004630.  
XX  
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX  
XX WPI; 2002-043418/06.  
XX  
XX New nucleotide sequence useful in the identification or Lactococcus  
PT lactis and related species.  
XX  
XX Claim 6; SEQ ID NO 1665; 2504pp; French.  
XX  
XX The present invention is related to a Lactococcus lactis nucleotide  
CC sequence (AB90521) and related proteins (AB93300-AB95621). The nucleic  
CC acid sequence is useful in the detection and/or amplification of nucleic  
CC acid sequence, particularly to identify Lactococcus lactis or related  
CC species. The proteins of the invention are useful for the biosynthesis or  
CC biodegradation of a composition of interest. The invention helps research  
CC in lactic bacteria, particularly useful in the production of yogurt and  
CC cheese. Note: The sequence data for this patent is based on equivalent  
CC patent WO20017334 (published 18-OCT-2001) which is available in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences. (Updated on 29-AUG-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 1072 AA:

Query Match 3.9%; Score 186.5; DB 5; Length 1072;  
Best Local Similarity 20.8%; Pred No. 0.00012;  
Matches 193; Conservative 136; Mismatches 363; Indels 235; Gaps 45;  
  
QY 1 MDSBSYYQ-NSTNQTPRSDVLDHRNQITN-DCAISDSEDELEKSELESEVVKSE 58  
Db : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
305 VASSESTQDANGASLYPISEASSVTDNTLNSISLSDSISSQTE-----NSQ 352  
  
QY 59 KQOHHOETSDNAKPLTRKSGSIIKKKSLTKDRTNPKSLSGDDTINSCHKRNYN 118  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
353 SGASSTAEISYSENSLSLSSNQINSNSN-SEKD--SNQSLG---SSMSGNESEHSN 406  
  
QY 119 MSLRKDPYLKONTDSTNNHTLAIPIPIPTIYNANKSRKKSQLENLPLIKK 178  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
407 SSNI-----NETNNSSEI-----TNILPPSNPTESNS-----VSDQ 437  
  
QY 179 TIGRNNNFENDLYSPMTKMTDSDITNTS-----TTANHKLGIG----- 222  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
438 TSSEASTNSNSLSLSSNSISDSSEATNSDFSNVAEVANNSLAVNNSSSVLSST 497  
  
QY 223 ---ATTLVG---GTQTATATATAAGRRPSSSIDSEADSHASRSQTEEDVCFPMVG 276  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
498 STADNLGINSQSDNLTKDSRISITSGAFLSNQTSSEASTNSNSISLSPSNISSTSV 557  
  
QY 277 DHRVNGIDPDEIDFR-----EEREAYLQKQMTAKNLRIDER--QNLKNN---T 325  
Db : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
558 ES--TTSSSNFNVAVANNSLAVNNSSSVLSSTSTADN-LEINQFGSONLTKDSSEIS 615  
  
QY 326 TSGASRHPYHHNNKNNKNGDGGGSSMAALKYTPKNIKLTLSRFEFTHENSSEEEI 385  
||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 616 TSCAFL-----SSNQTSSEASSNMSIN-SPSLSLLT-SNSESATNQSNSEAT 664  
QY 386 YELKTKQPPYVDQLSLTSTSTSGSGSQVFKGARGISDINGSLP--DRPSLPH 443  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
665 -----KVDNNSSTHSSNINLNSGNSDSDSDSD-SDSNLSSSPNLETNQITIS 712  
QY 444 SESEETIHAPDIPSLVSPGOSVRDLFRNGBETWMLDCTCTDSEKMLAKAFGIHPLTAE 503  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
713 SKPSEVNNISENPKKVSSNSVQE-----NSTDHEMSTNPKSSISSTPSTT 758  
QY 504 DIRMOETREKVELPKSYFYVCFHTFEADKSEDYL-----EPINVIYVVFHDLTFHESP 559  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
759 SSSQ-----KESQSNLNTTGIN-----NPITFNSS 787  
QY 560 ISHPANVRVRQLRDYVD---VSADMLCYALIDBI-TDGFAPVIHGIYEADAIEDAVF 615  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
788 SENSE-----ASILTSYNNSESSEETCCLYISNEAQRDNGSEISHSLP--SSNSNNVS 841  
QY 616 TARDTDFSSMLQIGESRRKVTMLRLLSGKADVIKFAKRCQBEANSSSGYQRYNLQ 675  
Db : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
842 SIQS---QATLESSEKSTNKRSSLSIINS-----TSHPQNEQNSNSDEVKSMNN 890  
QY 676 QOOQOAPPPPPPIITSPINSTLNLSLSTGCGVGVGGINFGPNPTGNNTNTNTTGTG 735  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
891 VES-----ILQQLNSISNKTMSLSLTSQKLSVITLPSKSKVTNEKNSENSTVS 939  
QY 736 -----SPSPPOOQOHHG-TT-----NKSF-PIPDARPRADIAL--YLGDITQDHIITMP 779  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
940 EEKLTKTPQKNDESQNLGQITALDLSFNKEVETMEDSKTVPDKVLNENGRSQNKKT-- 997  
QY 780 QNLLAYEK---IFSRSHSNYLAQLQVSPNSNNKITENPKSKITLGTMLVPLNLVTGLFG 836  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
998 -STIAKDNKVPKESSEFN---SKILSDNNILKTKVLKKHGLISD-----NLLIGLI- 1047  
QY 837 MNRVVPGEQTNLGHFPGIVGVLIPII 863  
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
1048 -----TILGACF-----IPFIPIL 1060  
  
Search completed: June 13, 2006, 14:08:12  
Job time : 207 secs



354 QY MAALKVTPKNTLKKTLSPBETHNSSSSEIYELTKQDPVKYDOLSLTSTST-- 411  
 DD MAALKTFTNLTLLKTLSPBETHNSSSSEIYELTKQDPVKYDOLSLTSTSTG 420  
 361 MAALKTFTNLTLLKTLSPBETHNSSSSEIYELTKQDPVKYDOLSLTSTSTG 420  
 QY SCGSGGVKFCGARISDKINGSLPDRSLPHSESBEHIAADPSLVSGOSVRDLPRN 471  
 DD SCGSGGVKFCGARISDKINGSLPDRSLPHSESBEHIAADPSLVSGOSVRDLPRN 480  
 472 GEETWMLDCTCPTDS EKXWLAKAFGLHPLTAEDIRHOETREKVLPKSYFVCFTHEAD 531  
 DD GEETWMLDCTCPTDS EKXWLAKAFGLHPLTAEDIRHOETREKVLPKSYFVCFTHEAD 540  
 532 KESSEDVLPIINVIVVFHDGLTFHPSPISHPANVRRVRQLRDYDVSAWMLCVALIDE 591  
 DD KESSEDVLPIINVIVVFHDGLTFHPSPISHPANVRRVRQLRDYDVSAWMLCVALIDE 600  
 592 ITDGFAPIVHGIEYEADAIDAVPTARDTDFSSMLQIGESRRRVMTLRLLSKGADVTK 651  
 DD ITDGFAPIVHGIEYEADAIDAVPTARDTDFSSMLQIGESRRRVMTLRLLSKGADVTK 660  
 601 ITDGFAPIVHGIEYEADAIDAVPTARDTDFSSMLQIGESRRRVMTLRLLSKGADVTK 660  
 QY MFAKRQCEANSSSGYQRYOYNLQQQQQQ--APPENPNPILTSPINSLTNLSLGTSTVG 709  
 DD MFAKRQCEANSSSGYQRYOYNLQQQQQQ--APPENPNPILTSPINSLTNLSLGTSTVG 720  
 710 GVGGEINFGNPNGNNTWTNTTSCSPSPQQQQQGRTWKNSPFPDARPDADIALYLGG 769  
 DD GVGGEINFGNPNGNNTWTNTTSCSPSPQQQQQGRTWKNSPFPDARPDADIALYLGG 779  
 770 DIQDHITTFQNLLAYEKIPSRSHSNYLAQOVESPNSNNKITMFSKTTLTGMTVLPLN 829  
 DD DIQDHITTFQNLLAYEKIPSRSHSNYLAQOVESPNSNNKITMFSKTTLTGMTVLPLN 839  
 830 LVTLGFANVRVPGEGNLTGMCFVGVLFIIGISPFIFAGMWLKLKLNNSIQEONGNR 889  
 DD LVTLGFANVRVPGEGNLTGMCFVGVLFIIGISPFIFAGMWLKLKLNNSIQEONGNR 899  
 890 PIPNHSSRRSIRSGLLKHGKNKSIIISPNNKYE 922  
 DD PIPNHSSRRSIRSGLLKHGKNKSIIISPNNKYE 932

RESULT 2  
 ID OSJAW3 CANAL PRELIMITARY; PRT; 929 AA.  
 AC OSJAW3:  
 DT 26-APR-2005, integrated into UniProtKB/TrEMBL.  
 DT 26-APR-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 6.  
 DE Hypothetical protein ALRI.  
 DE Name=ALRI; ORFNames=Cao19.1607;  
 DE Candida albicans SC5314.  
 DE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 DE Saccharomycetales; mitospotic Saccharomycetales; Candida.  
 CC NCBI\_TaxID=237561;  
 CC (1)  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RP STRAIN=SC5314;  
 RC PubMed=15123810; DOI=10.1073/pnas.0401648101;  
 RX Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S.,  
 RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,  
 RA Davis R.W., Scherer S.;  
 RT "The diploid genome sequence of Candida albicans.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334(2004).  
 RL CAUTION: The sequence shown here is derived from an  
 CL EMBL/GenBank/DDJB whole genome shotgun (WGS) entry which is  
 CC preliminary data.

---

Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
 Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; AAC001000013; EAL02896.1; -; Genomic\_DNA.  
 DR GO: 0016020; C:membrane; IEA.  
 DR GO: 0046873; P:metal ion transporter activity; IEA.  
 DR GO: 0030001; P:metal ion transport; IEA.  
 DR InterPro: IPR002523; Mg2+\_transptCorA.  
 DR Pfam: PF01544; CorA; 1.  
 DR Hypothetical protein.  
 SKW SEQUENCE 929 AA: 103372 MW: 98ADP462425595E4 CRC64;

OS Debaromyces hansenii (Yeast) (Torulaspora hansenii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaromyces.  
OX NCBI\_TaxID=4959;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=ATCC 36219 / CBS 767;  
RX PubMed=1529592; DOI=10.1038/nature02579;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
RA Boissarie A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicoud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron J., Scarpelli C., Gaillardin C., Weissenbach J.,  
RA Wincker P., Souciet J.-L.;  
RT \*Genome evolution in yeasts.\*;  
RL Nature 430:35-44(2004).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs license  
CC  
CC EMBL; CR382138; CAC89451.1; -; Genomic\_DNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0046873; F:metal ion transporter activity; IEA.  
DR GO; GO:0030001; P:metal ion transport; IEA.  
DR InterPro; IPR004523; Mg2+-transptCorA.  
DR Pfam; PF01544; CorA; 1.  
KW Complete proteome.  
SQ SEQUENCE 936 AA; 104547 MW; 8F050BA9EA942104 CRC64;  
  
Query Match 42.7%; Score 2056; DB 2; Length 936;  
Best Local Similarity 46.5%; Pred. No. 2.1e-108;  
Matches 485; Conservative 108; Mismatches 223; Indels 226; Gaps 31;  
  
QY 1 MSDSYQNSITNPIPR-----SDEVLDDHNRQITN-----DCAISDSEDELE 45  
DB 1 MSDSFSRFSRSLIHLRAANTGEVLNDRHQPSLPIRGTCGQMGQSSASVDDDDDDY 60  
QY 46 LKSELESEVVKKEKQHQHOB-----TSDNAKPLRKSSGSSKKSNLTKDKRIT 96  
DB 61 QSRSMQDAQYQSQSQSQPPHPOLQGSRSVDSQPPKAKQSQPSQYRS----- 111  
QY 97 NPMLSGGDDTINSCHNRNRYNNSLRKDPYLNKNTDQ-----STNNHTLAIPPIP 150  
DB 112 RSGCGMTGAPKQSRFRSQSFSGQAKSDQDEVRRHQSTAAAPPGGILMKSDNPPSK 154  
QY 151 IPTPIITNANKS-----RRKSQLENLPP-----LIR-----KKT 179  
DB 155 RSGCGMTGAPKQSRFRSQSFSGQAKSDQDEVRRHQSTAAAPPGGILMKSDNPPSK 214  
QY 180 IGRNNS---NNFENDLVSPMTOKTNDSEDTNTSTTANIMKLGITGATTLGVGTGTATA 236  
DB 215 SKRSNSMINDTLEADLAAPWTKSTNES-----MTSEAQNANK----- 252  
QY 237 TATAAAGRRPSRSTSDSEADSHASRSSQTEEDVCFPVGDHVRVNGIDFDEIFREE 296  
DB 253 -----RERRSRSTSDSEADSHASRSSQTEEDVCFPMLAEHVRVKGIDFDEIFRDE 307  
QY 297 RE-EAYL--OKQHIA-KNILRIDEFQNLKNTTSGASRHPYHHHNNKNGGDDGGS 352  
DB 308 KENEMHLKEEQQIAERTAMRD-----TGHGSARDV-----GP 341  
QY 353 SMAALKYTPKNILKTLRSFETH-----SSSEEEYE-----LKTQKPPYKY 398  
DB 342 SKNAKLYTPSNILGA-SPGKFRKNTYAGDMSSSGINISERAQSDSVTVDDSGARRY 400  
QY 399 DDQLSLTSTSTSGSGSGQVFGGARISDGINGGSLPDRFSLPHSSEETHAPDPSL 458  
DB 401 NEKI-LAGDDDISTE-----NVKFGGTRIND--NDSALPERFSPFCSESEETHAPDPSL 452  
QY 459 VSPGQSVRLPFRGETWMLDCTCPTDSEKMKLAKAFGIHLTAEDRMQETREKVELPK 518

DB 453 VKPGQSVSDLFRNGSGTWLDCVCPDAEMKMKVAKAFGIHLTAEDRMQETREKVELPK 512  
QY 519 SYVYFCFTEADKESDEYLEPINVIYVPHDGIILTPHSPISHPANVRVRVRLQRLDYD 578  
DB 513 NYVYFCFTEADKESDEYLEPINVIYVPHDGIILTPHSPISHPANVRVRVRLQRLDYD 572  
QY 579 VSADMLCVALIDEITDGFAPVHIGIEYEADAIEDAVPTARDTDFSSMLQRIQESRRKVT 638  
DB 573 VSADMLCVALIDEITDGFAPVHIGIEYEADAIEDSVFAARQVGFSTMIQRIQESRRKVT 632  
QY 639 LARLLSGKADVIKFAKRCOEBA-----NSSSGYV-QROYNLQOQQOQAPP 683  
DB 633 LARLLSGKADVIKFAKRCOEBAANQINIFQAHNNMNGMNYTSQTSNGASQSHTHSGS 692  
QY 684 PPNPIITSTINSTLNLSLSTGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGV 743  
DB 693 YTHGYNESPYQGYNNNS-----NPGYNNNSNPGYNNNSNPGYNNNSNPGY 743  
QY 744 QQ-----HGTTKNSF-----PIPD-ARPRADIALYLGDIQDHIITMFQNL 784  
DB 744 NOGGFVGTGPGVGGVNNMIDPQAFYKTKTQDPTQPRADIALYLGDIQDHIITMFQNL 803  
QY 785 YKIFSRSHENYLAQLQVESFNSNNKITEMFSKITLIGTMLVPLALVTGLFGNVVRPE 844  
DB 804 YKIFSRSHENYLAQLQVESFNSNNKITEMFSKITLIGTMLVPLALVTGLFGNVVRPE 863  
QY 845 GGTN--LGMFFGIVGLIPIITGSPFAQMWLKLNNISIEGQNNQNRPIFNHSSRSIRS 902  
DB 864 GDDHFGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGG 914  
QY 903 LGLKKH--GKNKSIISFNKYE 922  
DB 915 YGLKTRTKKAAKSIISFNKYE 936  
  
RESULT 4  
O6PV15\_CANGA PRELIMINARY; PRT; 803 AA.  
ID O6PV15;  
AC 19-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2004, sequence version 1.  
DE 07-FEB-2006, entry version 12.  
DS Similar to sp|Q08269|Saccharomyces cerevisiae YOL130w divalent cation transporter.  
GN OrderedLocusNames=CAGL0E01617g;  
OS Candida glabrata (Yeast) (Torulopsis glabrata).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5478;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=ATCC 2001 / CBS 138;  
RX PubMed=1529592; DOI=10.1038/nature02579;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
RA Boissarie A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicoud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron J., Scarpelli C., Gaillardin C., Weissenbach J.,  
RA Wincker P., Souciet J.-L.;  
RT \*Genome evolution in yeasts.\*;  
RL Nature 430:35-44(2004).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; CR380951; CAG58678.1; -; Genomic\_DNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0046873; F:metal ion transporter activity; IEA.  
DR GO; GO:0030001; P:metal ion transport; IEA.  
DR InterPro; IPR004523; Mg2+-transptCorA.  
DR Pfam; PF01544; CorA; 1.  
DR



Q08269	Q02811	008269	Q02811	integrated into UniProtKB/Swiss-Prot.
01-NOV-1997	sequence version 1.	01-NOV-1997	sequence version 1.	
07-MAR-2006	entry revision 41	07-MAR-2006	entry revision 41	
DT	Magnesium transporter ALR1 (Aluminum resistance protein 1).	DT	Magnesium transporter ALR1 (Aluminum resistance protein 1).	
DE	MagAlr1, OrderedLocustName:YOL130W; Saccharomyces cerevisiae (Baker's yeast).	DE	MagAlr1, OrderedLocustName:YOL130W; Saccharomyces cerevisiae (Baker's yeast).	
GN	Saccharomyces cerevisiae (Baker's yeast).	GN	Saccharomyces cerevisiae (Baker's yeast).	
OS	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.	OS	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.	
OC	NCBI_TaxID=4932	OC	NCBI_TaxID=4932	
OX	NCBI_TaxID=4932	OX	NCBI_TaxID=4932	
RN	NUCLEOTIDE SEQUENCE [GENOMIC DNA].	RN	NUCLEOTIDE SEQUENCE [GENOMIC DNA].	
RP	STRAIN=5288c / FV1679;	RP	STRAIN=5288c / FV1679;	
RC	MEDLINE=97051588; PubMed=8896265;	RC	MEDLINE=97051588; PubMed=8896265;	
RX	DOI=10.1092/STIC11037-0061(199609);12:108-1013;AID-YEA980>3.3.CO;2-X;	RX	DOI=10.1092/STIC11037-0061(199609);12:108-1013;AID-YEA980>3.3.CO;2-X;	
RA	Casamayor A., Khalid H., Balcells L., Aldea M., Casas C., Arino J.	RA	Casamayor A., Khalid H., Balcells L., Aldea M., Casas C., Arino J.	
RT	*Sequence analysis of a 13.4 kbp fragment from the left arm of chromosome XV reveals a malate dehydrogenase gene, a putative Ser/Thr protein kinase, the ribosomal L25 gene and four new open reading frames.	RT	*Sequence analysis of a 13.4 kbp fragment from the left arm of chromosome XV reveals a malate dehydrogenase gene, a putative Ser/Thr protein kinase, the ribosomal L25 gene and four new open reading frames.	
RL	Yeast 12:1013-1020(1996).	RL	Yeast 12:1013-1020(1996).	
RN	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].	RN	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].	
RC	STRAIN=5288c / FV1679;	RC	STRAIN=5288c / FV1679;	
RX	MEDLINE=9713270; PubMed=9169874;	RX	MEDLINE=9713270; PubMed=9169874;	
RA	Dujon B., Albermann K., Aldea M., Alexandraki D., Ansorge W., Arino J., Benes V., Bohn C., Bolotin-Fukuhara M., Bordonne R., Boyer J., Camasses A., Casamayor A., Casas C., Chert G., Cleplich C., Daiguen-Fornier B., Dang V.-D., de Haan M., Dellus H., Durand P., Fairhead C., Goffeau A., Gaillon L., Gallison F., Gamo F.-J., Gancedo C., Gellmann A., Goulding S.E., Grivell L.A., Habbig B., Hand N.J., Hani J., Hattenhorst U., Hebling U., Hernandez Y., Herrero E., Heumann K., Hiesel R., Hilger F., Hofmann B., Hollenber C.P., Hughes B., Jauniaux J.-C., Kalogeropoulos A., Katsoulou C., Kordes E., Leuven M.J., Landt O., Louis E.J., Maere A.-C., Madania A., Manhaupt G., Marck C., Martin R.P., Meves H.-W., Michaux G., Pannhuys V., Parle-McMormott A.G., Pearson B.M., Perrin A., Peterson B., Poch O., Pohl T.M., Poirey R., Portetelle D., Pujol A., Purnelle B., Ramezani Rad M., Rechmann S., Schwager C., Schweizer M., Sor F., Sterky F., Tarasov I.A., Teodoru C., Tettelin H., Thierry A., Tobiasch E., Tzermia M., Uhlen M., Unseld M., Valens M., Vandenbol M., Vetter I., Vitek C., Voet M., Voicard G., Voss H., Wambolt R., Wedter H., Wiemann S., Winsor B., Wolfe K.H., Zolner A., Zumstein E., Kline K.	RA	Dujon B., Albermann K., Aldea M., Alexandraki D., Ansorge W., Arino J., Benes V., Bohn C., Bolotin-Fukuhara M., Bordonne R., Boyer J., Camasses A., Casamayor A., Casas C., Chert G., Cleplich C., Daiguen-Fornier B., Dang V.-D., de Haan M., Dellus H., Durand P., Fairhead C., Goffeau A., Gaillon L., Gallison F., Gamo F.-J., Gancedo C., Gellmann A., Goulding S.E., Grivell L.A., Habbig B., Hand N.J., Hani J., Hattenhorst U., Hebling U., Hernandez Y., Herrero E., Heumann K., Hiesel R., Hilger F., Hofmann B., Hollenber C.P., Hughes B., Jauniaux J.-C., Kalogeropoulos A., Katsoulou C., Kordes E., Leuven M.J., Landt O., Louis E.J., Maere A.-C., Madania A., Manhaupt G., Marck C., Martin R.P., Meves H.-W., Michaux G., Pannhuys V., Parle-McMormott A.G., Pearson B.M., Perrin A., Peterson B., Poch O., Pohl T.M., Poirey R., Portetelle D., Pujol A., Purnelle B., Ramezani Rad M., Rechmann S., Schwager C., Schweizer M., Sor F., Sterky F., Tarasov I.A., Teodoru C., Tettelin H., Thierry A., Tobiasch E., Tzermia M., Uhlen M., Unseld M., Valens M., Vandenbol M., Vetter I., Vitek C., Voet M., Voicard G., Voss H., Wambolt R., Wedter H., Wiemann S., Winsor B., Wolfe K.H., Zolner A., Zumstein E., Kline K.	
RT	*The nucleotide sequence of Saccharomyces cerevisiae chromosome XV.	RT	*The nucleotide sequence of Saccharomyces cerevisiae chromosome XV.	
RL	Nature 367:98-102(1997).	RL	Nature 367:98-102(1997).	
RN	NUCLEOTIDE SEQUENCE [GENOMIC DNA].	RN	NUCLEOTIDE SEQUENCE [GENOMIC DNA].	
RC	STRAIN=5288c; Refs A., Richardson A., Kane M., Bequi M., Teycher E., Hu F., Vannberg P., Weger J., Kramer J., Melita D., Kelley P., A., Zuo D., Raphael J., Hogg C., Simpson D., Williams J., Cantley A., Gonzalez J., Vasconcelos A.T., Simpson A., Kolodner R., Harlow E., Labeat J.	RC	STRAIN=5288c; Refs A., Richardson A., Kane M., Bequi M., Teycher E., Hu F., Vannberg P., Weger J., Kramer J., Melita D., Kelley P., A., Zuo D., Raphael J., Hogg C., Simpson D., Williams J., Cantley A., Gonzalez J., Vasconcelos A.T., Simpson A., Kolodner R., Harlow E., Labeat J.	
RT	*Creation of the YLEP clone resource: cloning of Saccharomyces cerevisiae ORFs in the Gateway recombinational cloning system.	RT	*Creation of the YLEP clone resource: cloning of Saccharomyces cerevisiae ORFs in the Gateway recombinational cloning system.	
RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.	RL	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.	
RN	FUNCTION	RN	FUNCTION	
RP	PubMed=9430719; DOI=10.1074/jbc.273.3.1727;	RP	PubMed=9430719; DOI=10.1074/jbc.273.3.1727;	
RA	MacDiarmid C.W., Gardner R.C.	RA	MacDiarmid C.W., Gardner R.C.	
RT	*Overexpression of the Saccharomyces cerevisiae magnesium transport system confers resistance to aluminum ion.	RT	*Overexpression of the Saccharomyces cerevisiae magnesium transport system confers resistance to aluminum ion.	
RL	J. Biol. Chem. 273:1727-1732(1998).	RL	J. Biol. Chem. 273:1727-1732(1998).	
CC	!- FUNCTION: Plasma membrane magnesium transporter.	CC	!- FUNCTION: Plasma membrane magnesium transporter.	
CC	!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (Potential).	CC	!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (Potential).	
CC	!- SIMILARITY: Belongs to the corA family.	CC	!- SIMILARITY: Belongs to the corA family.	
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms	CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms	
CC	Distributed under the Creative Commons Attribution-NoDerivs License	CC	Distributed under the Creative Commons Attribution-NoDerivs License	
CC	EMBL; UA1293; AAC49462.1; -; Genomic DNA.	CC	EMBL; UA1293; AAC49462.1; -; Genomic DNA.	
DR	EMBL; 274872; CA99150.1; -; Genomic DNA.	DR	EMBL; 274872; CA99150.1; -; Genomic DNA.	
DR	EMBL; 274871; CA99149.1; -; Genomic DNA.	DR	EMBL; 274871; CA99149.1; -; Genomic DNA.	
DR	EMBL; AY692762; AAT92781.1; -; Genomic DNA.	DR	EMBL; AY692762; AAT92781.1; -; Genomic DNA.	
DR	PIR; S66827; S66827.	DR	PIR; S66827; S66827.	
DR	GeneOnline; 143552; -.	DR	GeneOnline; 143552; -.	

Ensembl: YOL130W; Saccharomyces cerevisiae.  
GenomeReviews: Y13140.GR; YOL130W.  
SGD; 500005490; ALR1.  
BioCyc: SCER-528-01:SCER-528-01-005413-MONOMER; --.  
LinkHub; Q08269; --.  
GO; GO:0005886; C:plasma membrane; IMP.  
GO; GO:0015082; F:di-, tri-valent inorganic cation transporte. .; IMP.  
GO; GO:0015693; P:magnesium ion transport; IDA.  
InterPro; IPR002523; Mg2+-transptCorA.  
Pfam; PF01544; CorA; 1.  
Complete proteome; Magnesium; Membrane; Transmembrane; Transport.  
CHAIN 1  
TRANSMEM 744 764 Potential.  
TRANSMEM 773 793 Potential.  
CONFLICT 13 13 N -> Y (in Ref. 1).  
SEQUENCE 859 AA; 9586 MW; 6DA44CA708FA2693 CRC64;  
Query Match 34.9%; Score 1680; DB 1; Length 859;  
Best Local Similarity 41.5%; Pred. No. 5e-87;  
Matches 405; Conservative 137; Mismatches 258; Indels 176; Gaps 31;  
QY 1 MSDESYYQNSTTNOIPRSDEVLDDHNRQITNDCAISDEDELEL---KSELESEVKS 57  
DB 1 MSSSSS---SSSESNLSRS-----NSLANTWVSMKTEHTGLYDHRQHPDLSLVRH 49  
QY 58 EKQOQHQETSDNAKPLTRKSGSIKKK-----NLTKDRITNPMSLSGDGDITNSGH 112  
DB 50 QPPTLNKKEIAKSTKPSIPKEQKSAATNSHVDVGSVPSGRMDPEDEGQMDVTV-AHI 108  
QY 113 KNR-----NTNMSSL-----RKDFYKLDNTDNTNNHT---HLAIPIPITPII 156  
DB 109 QLRASAILTSNARPSLAHSMHPORQLYVESNI-----HTPPKDVGVKRVYTMSSSTA 161  
QY 157 TWANKSRKQLENPLPIKKKTTIGRNNNNFNDLVSPMTKMTNDSIEDTNT----- 210  
DB 162 SSGNKS-KLSASSASPIKTVR-----KSLVSPVLBIHPHESKDSHKLAKPKK 210  
QY 211 ----STTANMKLGICATLGVGTGTATA---TATAAAGRPSRSSIDSAHSAHRSQQ 264  
DB 211 RTVYTSYSAHSSIN-PAVLLTKTSOKSDADDTLERKPVMTNTRASFDSV-SQASRDSQ 268  
QY 265 ETEEDVCFPMVDG-HRVNGIDFDEIFREEREAYLQKQIAK-NILRIDEPQNLK 322  
DB 269 ETEEDVCFPMQLHTRVNGIDFDELEAYAQFANAE---KSQFLASLQVFNQKYSNVSQ 325  
QY 323 -----NNTTSGASRHPYHHNNKNNKNGGGGSSMAALKYTPKMLKTLRSFRFTH 376  
DB 326 DIGPSTSTSGSS-----AALKYTPR-VSQGKESESTN 359  
QY 377 ENSSSEIEYELTKQPPYKDDQLSLTSTSGSGSGQVFKGG-ARISDINGGSL 435  
DB 360 ET-----EIEHKKDEHEKIKPSLHPGISFGKNKVEGENEIPNDPAYCSYQGTDFQI 414  
QY 436 PDRFLPFSSEETHAPDIPSLVSPQSVROLFRNGEETWMLDCTPTDSEMKMLAKAF 495  
DB 415 PNRPSFCSSEDETVHASDIPSLVSEGOFTFELFRGGEPTWMLDCSPTDDMRICAKAF 474  
QY 496 GHPIUADIEMQETREKVELPKSYVCFPTFEADKESDYLEPINVIYVFDHGLTF 555  
DB 475 GHPIUADIEMQETREKVELPKSYVCFPTFEADKESDYLEPINVIYVCRSGVLTFF 534  
QY 556 HPSPIHPANVRVRQLRDYVDVSDMLCYALIDEITDGFAPVTHIEYEADEADAVF 615  
DB 535 HFGPISHCANVRVRQLRDYVNVNSDMLCYALIDITDTSFAPVIOQIIEYEADEADSVF 594  
QY 616 TARDYDFSSMLQIGESRRKVTMLRLLSGKADVIKFAKQOEANSSSGYQRYQNLQ 675  
DB 595 MARDMDFAAMLQIGESRRKVTMLRLLSGKADVIKFAKQOEANGI-----643  
QY 676 QQQQQAQPPPPPIITSPINSTLNLSLSTGGGVGVGGINFGPNPTGNNTWNTWTYTG 735  
DB 644 -----GPAITSQIN-IANIQAQDNA-----SHIKNNSSTVTPNNY 678  
QY 736 SPSPQQQQQHGTWKSFPIDARPRADIALYGLDIDHITWQNLAYEKFGRSHN 795  
DB 679 APTTSQ-----PRGDIALYGLDIDHLLTWFQNLAYEKFGRSHN 720  
QY 796 YLAQQLQVESFNSNNKITEMFSKITLIGTMLVPLNLTGLFGMNVKVPCEGNTLQWFFGI 855

721 YLAQQLQVESFNSNNKITEMFSKITLIGTMLVPLNLTGLFGMNVKVPCE-NSSIAWFFGI 779  
QY 856 VGVLFIFIGSFIFAQWMLKLNNSI---EQQNGNRPFEF-----HSSRRSTRSLGKK 907  
DB 780 LGVLLLVLLVGLWFLASVWIKRIDPPATLNEAESAAGSAKSVISSFLPKRKRFRNDRSKNIV 839  
QY 908 HGG--NKSIISFPNNKY 921  
DB 840 RAGPSNKSVASLPSRY 855  
RESULT 6  
ID ALR2\_YEAST  
AC P43553;  
DT 01-NOV-1995, integrated into UniProtKB/Swiss-Prot.  
DT 07-MAR-2006, entry version 40.  
DE Magnesium transporter ALR2 (Aluminum resistance protein 2).  
GN Name=ALR2; OrderedLocusNames=YFL050C;  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=S288c / AB972;  
RX MEDLINE=95400292; PubMed=7670463;  
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
RA Sasanuma S., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
RA Yamazaki M., Tashiro H., Eki T.;  
RT \*Analysis of the nucleotide sequence of chromosome VI from  
RT Saccharomyces cerevisiae.\*;  
RL Nat. Genet. 10:261-268(1995).  
RN [2]  
RP FUNCTION.  
RX PubMed=9430719; DOI=10.1074/jbc.273.3.1727;  
RA MacDiarmid C.W., Gardner R.C.;  
RT \*Overexpression of the Saccharomyces cerevisiae magnesium transport  
RT system confers resistance to aluminum ion.\*;  
RL J. Biol. Chem. 273:1727-1732(1998).  
RN [3]  
RP LEVEL OF PROTEIN EXPRESSION.  
RX MEDLINE=22923965; PubMed=14562106; DOI=10.1038/nature02046;  
RA Ghacemaghani S., Ruh W.-K., Bower K., Howson R.W., Belle A.,  
RA Dephourse N., O'Shea E.K., Weisman J.S.;  
RT \*Global analysis of protein expression in yeast.\*;  
RL Nature 425:737-741(2003).  
CC --!- FUNCTION: Plasma membrane magnesium transporter.  
CC --!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein  
CC --!- MISCELLANEOUS: Present with 2130 molecules/cell.  
CC --!- SIMILARITY: Belongs to the corA family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; D50617; BAA09191.1; -; Genomic DNA.  
CC PIR; S56205; S56205.  
CC GerMOnline; 140106; --.  
CC Ensembl; YFL050C; Saccharomyces cerevisiae.  
CC GenomeReviews; D50617.GR; YFL050C.  
CC SGD; 500001844; ALR2.  
CC BioCyc; SCER-528-01:SCER-528-01-001921-MONOMER; --.  
CC LinkHub; P43553; --.  
CC GO; GO:0005886; C:plasma membrane; IMP.  
CC GO; GO:0015082; F:di-, tri-valent inorganic cation transporte. .; IMP.  
CC GO; GO:0015693; P:magnesium ion transport; IDA.  
CC InterPro; IPR002523; Mg2+-transptCorA.  
CC Pfam; PF01544; CorA; 1.  
CC Complete proteome; Magnesium; Membrane; Transmembrane; Transport.  
FW CHAIN 1 858 Magnesium transporter ALR2.  
FT /FTID=PRO\_0000201536.  
FT  
FT TRANSMEM 742 762 Potential.  
FT TRANSMEM 772 792 Potential.  
FT SEQUENCE 858 AA; 9668 MW; 7AAAFEC6086FF0618 CRC64;  
SQ  
Query Match 34.6%; Score 1666; DB 1; Length 858;



[illegible]

## RESULT 10

Query Match	24.38;	Score 1173;	DB 2;	Length 803;
Best Local Similarity	35.38;	Pred. No. 3.5e-58;		

Matches 310; Conservative 100; Mismatches 247; Indels 220; Gaps 22;

QY 41 EDELEL-KSELESEVVKSEKQOQHQBHT-----SDNAK 73  
 DB 1 ESEIELHESVMPQPKKKRRRRKASRKIPATOSSADMDTLHVCPTGSCSDSKK 150  
 QY 74 PLTRKS--GSSIK--KKSNUITDKRINTNPMISLGGDDITNSGKNRYN-----MSSLARK 124  
 DB 151 POSNKKHRRGRVKSFKSTLEVPDIPDIALKASMGSS--SSQHASR--YNEGKRGSDFTLGT 208  
 QY 125 DYFL-KONTDNTNNTHLAIPPIPIPTIITNANKSRKSOLENLPLPKKTKTGRN 183  
 DB 209 HYHLSKSRKSSSDSLYSVSSSIKNDNSDLSLSSSSSDVSGSDENLP--IDKTLVLY 266  
 QY 184 NSNNFENDLVPMTKMTNDSDINTSTTANMKKLIGAGATLGV-GTGTATATATAA 242  
 DB 267 EDPSPVHPRPSPATKSCAAVDCPHTTIPKPPQSDTDLTLPKTAQFSDFTTVPQ 326  
 QY 243 GRPSSRSI-----DSEADSHASRSQSTEDVCFPMVGDHIRVINGIDIDEIDFIRERE 298  
 DB 327 RLKSTHSTADNEDREVDSQVDNTRVVEEDVCFPMQEEESHVNGKIDFDELINFAEE-- 384  
 QY 299 EAYLQKMTAKNILRAIDFQNLKNNKNTTSGASRH-----PYHHSSNNKNNKGGGGSS 353  
 DB 385 ----LQKQ-----RQNTDHFRRSQVSTCKPFPFHWNDLSHPDND----- 420  
 QY 354 MAALKYTPKNILKTLRPEFTHENSSESSEIYELKTKQPPKYDDQLSLTSTSTSG 413  
 DB 421 -----PSSSL-----HSNNAEAAEVPLRSS--YSGGRRSVTSMDS---- 456  
 QY 414 SGSGQVFGGARISDGINGSGSLPDRFSLPHSESEETHAPDIPSLVSPG-QSVRDLPFRG 472  
 DB 457 -----PYRFSFFSSNQNETHASVLSLDLDDGATSFSLFCPE 494  
 QY 473 EETWMLDCTCPTDSEMKLAKAFGIHPLTAEDIRMOETREKVELFKSYFVCFHTFEADK 532  
 DB 495 KGVWMLDCLDPTDIEMRLSKAFSIHPLTTEDIRVQAREKVELFGSYFVCFASFQDP 554  
 QY 533 ESEDLPEINVIYVFDGILTFHSPISHPANVRVRQLRDYVVSADMLCYALIDEI 592  
 DB 555 ELANYLPLMIYVFRGLLTFHESATHPASVRRARQLRDYVVSADMLCYALIDDI 614  
 QY 593 TDGPAVHIGIEYEAIDAVFTARTDPSFSLQRIEGRSRKVTMLRLSLGKADYRM 652  
 DB 615 TDAFVPLIRIGIETEAIEDSVLVRSESDSMLRIGBCRKTKMGHFRLLYKADYRM 674  
 QY 653 FAKRCOEANSSGGYQRYNLOQOQOQAPPPPPPIITSPINSLNLSLGTSTGGGVG 712  
 DB 675 LAKRCNKWTIA-----PTG----- 686  
 QY 713 VGGINFGPNPTGNNTNTNTTGTSPSPPOQOQOQHITNKSFPIDARPRADIALYLGDIQ 772  
 DB 687 -----PTG-----EIGLYLGDIO 699  
 QY 773 DRIITMFQNLAYEKIFSRSHSNYLAQVFSNKNKITEMFSKITLIGTMLVPLNLT 832  
 DB 700 DILVMTSNLSQFEKILSRTHSNYLAQVFSNKNKITEMFSKITLIGTMLVPLNLT 759  
 QY 833 GLFGNVRVPGSGGTNLGMPFGIVGLI-FIIIGSFI 868  
 DB 760 GLFGNVPVPGSDTNLNLAWPFGILGVLLGSIAIGWII 796

RESULT 11  
 Q7LWD8 SCHPO PRELIMINARY; PRG: 499 AA.  
 ID Q7LWD8 SCHPO  
 AC Q7LWD8  
 DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 05-JUL-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 8.  
 DE Probable membrane protein YOL130w (Fragment).  
 GN Name=pi066;  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.

STRAIN=972 h-;  
 MEDLINE=20089027; PubMed=10620777;  
 DOI=10.1007/s10397-0061(20000115)16:1<71::AID-YEA505>3.0.CO;2-5;  
 Machida M., Yamazaki S., Kunihiro S., Tanaka T., Kushiida N., Jinno K.,  
 Haikawa Y., Yamazaki J., Yamamoto S., Sekine M., Oguchi A., Nagai Y.,  
 Sakai M., Aoki K., Ogura K., Kudoh Y., Kikuchi H., Zhang M.Q.,  
 Yanagida M.;  
 \*A 38 kb segment containing the cdc2 gene from the left arm of fission  
 yeast chromosome II: sequence analysis and characterization of the  
 genomic DNA and cDNAs encoded on the segment.\*;  
 Yeast 16:71-80(2000).  
 (2)  
 NUCLEOTIDE SEQUENCE.  
 STRAIN=972 h-;  
 RC Director-General of Biotechnology Center, Kushiida N., Machida M.;  
 Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
 CC Distributed under the Creative Commons Attribution-NoDerivs license  
 CC EMBL; AB04539; BAA21448.1; -; Genomic\_DNA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0046873; F:metal ion transporter activity; IEA.  
 DR GO; GO:0030001; P:metal ion transport; IEA.  
 DR InterPro; IPR002523; Mg2+-transpCorA.  
 DR Pfam; PF01544; CorA; 1.  
 DR NON\_TER  
 FT SEQUENCE 499 AA; 56728 MW; BGC371589B6D2C4C CRC64;  
 SQ  
 Query Match 23.7%; Score 1141.5; DB 2; Length 499;  
 Best Local Similarity 40.9%; Pred. No. 1.2e-56;  
 Matches 267; Conservative 60; Mismatches 143; Indels 183; Gaps 13;  
 QY 232 TTATA-----TATAAGRRPSRSI-----DSEADSHASRSQSTEDVCFPMVGDHIRVN 282  
 DB 7 TKSTAQFSDFTTVPQPKSTHSSADNEDREVDSQVDNTRVVEEDVCFPMQESSHVNK 66  
 QY 283 GDIDPDEIDFIREEREAYLQKMTAKNILRAIDFQNLKNNKNTTSGASRH-----PYHHH 337  
 DB 67 GDIDPDELNFAEE-----LQKQ-----RQNTDHFRRSQVSTCKPFP 105  
 QY 338 SNNKNNKNGDGGSGMAALKYTPKNILKTLRPEFTHENSSESSEIYELKTKQPPYK 397  
 DB 106 WNDLSHPDND-----PSSSL-----HSNNAEAAEVPLRSS--YY 139  
 QY 398 YDQLSLTSTSTSTSGSGQVKGFGARISDGINGSLPDRFSLPHSESEETHAPDIPS 457  
 DB 140 SGRRSVTSMSD-----PYRFSFFSSNQNETHASVLSLD 174  
 QY 458 LVSPG-QSVRDLPFRNGBETWMLDCTCPTDSEMKLAKAFGIHPLTAEDIRMOETREKVEL 516  
 DB 175 LDDGATSFSLFCPEKGVWMLDCLDPTDIEMRLSKAFSIHPLTTEDIRVQAREKVEL 234  
 QY 517 FKSYYFVCFHTPEADKSESDYLEPINVIYVFDGILTFHSPISHPANVRVRQLRDY 576  
 DB 235 FGSYYFVCFRSPQEPBLANYLEPLNMIYVFRGLLTFHESATHPASVRRARQLRDY 294  
 QY 577 VDSADMLCYALIDEITDGPAPVHIGIEYEAIDAVFTARTDPSFSLQRIEGRSRK 636  
 DB 295 VIVSDMLCYALIDITDAFVPLIRIGIETEAIEDSVLVRSESDSMLRIGRCKRT 354  
 QY 637 MTLRLLSGKADYIKMFAKRCOEANSSGGYQRYNLOQOQOQAPPPPPPIITSPINS 696  
 DB 355 MGFPELLYKADYIKMLAKRCNKWTIA----- 382  
 QY 697 TLNLSLGTSTGGVGGVGINFGPNPTGNNTNTNTTGTSPSPPOQOQOQHITNKSFP 756  
 DB 383 -----PTG----- 385  
 QY 757 DARPRADIALYLGDIQDRIITMFQNLAYEKIFSRSHSNYLAQVFSNKNKITEMFS 816  
 DB 386 -----EIGLYLGDIOHVLVMTSNLSQFEKILSRTHSNYLAQVFSNKNKITEMFS 439  
 QY 817 KITLIGTMLVPLNLTGLFGNVRVPGSGGTNLGMPFGIVGLI-FIIIGSFI 868  
 DB 440 KITLIGTMLVPLNLTGLFGNVPVPGSDTNLNLAWPFGILGVLLGSIAIGWII 492

RESULT 12

## RESULT 13

Query Match 22.88: Score 1099: DB 2: Length 663:

Matches	291:	Conservative	96:	Mismatches	245:	Indels	260:	Gaps	21:
---------	------	--------------	-----	------------	------	--------	------	------	-----

1. **Introduction**

2. **Background**

3. **Methodology**

4. **Results**

5. **Discussion**

6. **Conclusion**

7. **References**

8. **Appendix**

9. **Tables**

10. **Figures**

11. **Supplementary Materials**

12. **Notes**

13. **Abbreviations**

14. **Conflicts of Interest**

15. **Acknowledgments**

16. **Author Contributions**

17. **Funding**

18. **Data Availability Statement**

19. **References**

20. **Appendix**

21. **Tables**

22. **Figures**

23. **Supplementary Materials**

24. **Notes**

25. **Abbreviations**

26. **Conflicts of Interest**

27. **Acknowledgments**

28. **Author Contributions**

29. **Funding**

30. **Data Availability Statement**

31. **References**

32. **Appendix**

33. **Tables**

34. **Figures**

35. **Supplementary Materials**

36. **Notes**

37. **Abbreviations**

38. **Conflicts of Interest**

39. **Acknowledgments**

40. **Author Contributions**

41. **Funding**

42. **Data Availability Statement**

43. **References**

44. **Appendix**

45. **Tables**

46. **Figures**

47. **Supplementary Materials**

48. **Notes**

49. **Abbreviations**

50. **Conflicts of Interest**

51. **Acknowledgments**

52. **Author Contributions**

53. **Funding**

54. **Data Availability Statement**

55. **References**

56. **Appendix**

57. **Tables**

58. **Figures**

59. **Supplementary Materials**

60. **Notes**

61. **Abbreviations**

62. **Conflicts of Interest**

63. **Acknowledgments**

64. **Author Contributions**

65. **Funding**

66. **Data Availability Statement**

67. **References**

68. **Appendix**

69. **Tables**

70. **Figures**

71. **Supplementary Materials**

72. **Notes**

73. **Abbreviations**

74. **Conflicts of Interest**

75. **Acknowledgments**

76. **Author Contributions**

77. **Funding**

78. **Data Availability Statement**

79. **References**

80. **Appendix**

81. **Tables**

82. **Figures**

83. **Supplementary Materials**

84. **Notes**

85. **Abbreviations**

86. **Conflicts of Interest**

87. **Acknowledgments**

88. **Author Contributions**

89. **Funding**

90. **Data Availability Statement**

91. **References**

92. **Appendix**

93. **Tables**

94. **Figures**

95. **Supplementary Materials**

96. **Notes**

97. **Abbreviations**

98. **Conflicts of Interest**

99. **Acknowledgments**

100. **Author Contributions**

101. **Funding**

102. **Data Availability Statement**

103. **References**

104. **Appendix**

105. **Tables**

106. **Figures**

107. **Supplementary Materials**

108. **Notes**

109. **Abbreviations**

110. **Conflicts of Interest**

111. **Acknowledgments**

112. **Author Contributions**

113. **Funding**

114. **Data Availability Statement**

115. **References**

116. **Appendix**

117. **Tables**

118. **Figures**

119. **Supplementary Materials**

120. **Notes**

121. **Abbreviations**

122. **Conflicts of Interest**

123. **Acknowledgments**

124. **Author Contributions**

125. **Funding**

126. **Data Availability Statement**

127. **References**

128. **Appendix**

129. **Tables**

130. **Figures**

131. **Supplementary Materials**

132. **Notes**

133. **Abbreviations**

134. **Conflicts of Interest**

135. **Acknowledgments**

136. **Author Contributions**

137. **Funding**

138. **Data Availability Statement**

139. **References**

140. **Appendix**

141. **Tables**

142. **Figures**

143. **Supplementary Materials**

144. **Notes**

145. **Abbreviations**

146. **Conflicts of Interest**

147. **Acknowledgments**

148. **Author Contributions**

149. **Funding**

150. **Data Availability Statement**

151. **References**

152. **Appendix**

153. **Tables**

154. **Figures**

155. **Supplementary Materials**

156. **Notes**

157. **Abbreviations**

158. **Conflicts of Interest**

159. **Acknowledgments**

160. **Author Contributions**

161. **Funding**

162. **Data Availability Statement**

163. **References**

164. **Appendix**

165. **Tables**

166. **Figures**

167. **Supplementary Materials**

168. **Notes**

169. **Abbreviations**

170. **Conflicts of Interest**

171. **Acknowledgments**

172. **Author Contributions**

173. **Funding**

174. **Data Availability Statement**

175. **References**

176. **Appendix**

177. **Tables**

178. **Figures**

179. **Supplementary Materials**

180. **Notes**

181. **Abbreviations**

182. **Conflicts of Interest**

183. **Acknowledgments**

184. **Author Contributions**

185. **Funding**

186. **Data Availability Statement**

187. **References**

188. **Appendix**

189. **Tables**

190. **Figures**

191. **Supplementary Materials**

192. **Notes**

193. **Abbreviations**

194. **Conflicts of Interest**

195. **Acknowledgments**

196. **Author Contributions**

197. **Funding**

198. **Data Availability Statement**

199. **References**

200. **Appendix**

201. **Tables**

202. **Figures**

203. **Supplementary Materials**

204. **Notes**

205. **Abbreviations**

206. **Conflicts of Interest**

207. **Acknowledgments**

208. **Author Contributions**

209. **Funding**

210. **Data Availability Statement**

211. **References**

212. **Appendix**

213. **Tables**

214. **Figures**

215. **Supplementary Materials**

216. **Notes**

217. **Abbreviations**

218. **Conflicts of Interest**

219. **Acknowledgments**

220. **Author Contributions**

221. **Funding**

222. **Data Availability Statement**

223. **References**

224. **Appendix**

225. **Tables**

226. **Figures**

227. **Supplementary Materials**

228. **Notes**

229. **Abbreviations**

230. **Conflicts of Interest**

231. **Acknowledgments**

232. **Author Contributions**

233. **Funding**

234. **Data Availability Statement**

235. **References**

236. **Appendix**

237. **Tables**

238. **Figures**

239. **Supplementary Materials**

240. **Notes**

241. **Abbreviations**

242. **Conflicts of Interest**

243. **Acknowledgments**

244. **Author Contributions**

245. **Funding**

246. **Data Availability Statement**

247. **References**

248. **Appendix**

249. **Tables**

250. **Figures**

251. **Supplementary Materials**

252. **Notes**

253. **Abbreviations**

254. **Conflicts of Interest**

255. **Acknowledgments**

256. **Author Contributions**

257. **Funding**

258. **Data Availability Statement**

259. **References**

260. **Appendix**

261. **Tables**

262. **Figures**</

4 SESYQNSTNQIPRSEVLDHHRNQITNDCAISDSEDELELKSELEVVKSEKQQH 63

7 SEKYNCLTRFSTPVP-----LDDHRFQL-----DSPPRIATADI-SLSRQNTAQHAY 54

[illegible]

04 AQE1SDNAKFLINKSGSSIAKKNELIDRKRIINFMELSGDDIINSGAARKRNINMSSLR 123

55 HOETPO-----BPPI.I.STONAI.BEAGSI.SPDFOATID 87

**THE UNIVERSITY OF CHICAGO**

124 KDFYKD-----NTDDNSTNNH-----THLAIPPIPTPIITNANKSRRK--SQ 167

[illegible]

88 DDRSGKDINTLGRFSDPNGNVRHGR TWSRTHQEL-----ANMSRESSPSA 134

[illegible]

168 L E N L P P L I K K K T I G R N N S N F E N D L V S P M I R K R I N D S E D I T N T S T A N H M K L G I G A T T L G 227

[illegible]

-----NOV EAT ADFNNRNNNT DESJNATE FDDER-----IDQKIVS 173

228 VGTGTTATATAAGRRPSRSSIDSEADSHASRSSOETEEDVCFPMVGDHIRVNGIDFD 287

[illegible]

176 GGTHPRRPTFSNAA-IRPQPGDIQLEPN-----DESCVPTVEQGRIPVIDYE 223

288 EIDFIREEREAYLQKQMIKNIILRIDEFQNLKNNYTSGARHP---YHHHSNNKKN 344

[illegible][illegible]

QO: CO-0030001; P-metal ion transport; IEA.  
 InterPro: IPR001356; Homeobox.  
 InterPro: IPR002523; Mga\_2\_transpCorA.  
 InterPro: IPR01544; CorA\_1.  
 DR PROSITE: PS00027; HOMEBOX\_1; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 460 AA: 52281 MW: 5F04F600DFD76400 CRC64;  
  
 Query Match 21.7%; Score 1044; DB 2; Length 460;  
 Best Local Similarity 42.5%; Pred. No. 3.8e-51;  
 Matches 202; Conservative 76; Mismatches 93; Indels 104; Gaps 3;  
  
 QY 437 DRSLPSESEETIHADIPSLVPSQGSVRDLFRNGEETWLLDCTCPSEMKMLAKAGC 496  
 DB 77 DHEAFFSECEETIRANFSEIFKDCAKELFDAHKGTYWLLDCTPDAENKLLAKAGC 136  
  
 QY 497 IHLPTAEDIRMOETREKVELPKSYFYFCFHTPEADKSESDYLEPNIYVIVPHDQILYTH 556  
 DB 137 IHLPTVEDIQHREAREKVEFRNYFYFAFVAHQVESDDFEMVPPYLCVPEGVISFH 196  
  
 QY 557 FSPIGHANVRVRQLRDYVDSADMLCYALIDITDGFAPVINGIEYEAADATEDAVFT 616  
 DB 197 YSVPHCATVKKRIQLKDHVTTPDMTCYVAVIDDTSFAPVIREIEQEGSIEVQTVFD 256  
  
 QY 617 ARUTDFSSMLQRTGESRKVYNTLMRLASGADVINKFAKCOEEANSSGYSYQRYNQVLAQ 676  
 DB 257 ARDGFENMLRRIGAKAKTSLMRLASGADVGMFTKRVSE----- 299  
  
 QY 677 QQQQAPPNPPIITSPINSTLNLNLSGTSGGGVGGVFNFGPNFGNTWTNTNTVGS 736  
 DB 300 ----- 299  
  
 QY 737 PSPPQQQQHGHITNKSFPIDAPRADIALYLGDIQDIHTYFQMLLAYEKIFGRSHSY 796  
 DB 300 -----GLSDHVPKHSILYLGDIQDIHTYFQMLLAYEKILSRSHLY 342  
  
 QY 797 LAQLOVESPNNSKITEMFKITLGLTMVPLNLVTLGLFQNVVRVPGEGT-----NLGMF 852  
 DB 343 LSLQGVQSIDATHVTDTGLKIVTIGTILIPNNVTLGLFQNVVRVPGEGT-----NLGMF 402  
  
 QY 853 FGTVGVLPITIGSFPAQMWLKLNNISIEQGNQNRPIFNHSSRRSIRSLGLK 907  
 DB 403 FGILGVGLCIIVTATVAKVLAHVKKINGN-----FNSASDINSIRMSKR 451  
  
 RESULT 14  
 ID QAPHQ2.USTMA PRELIMINARY; PRT: 881 AA.  
 AC QAPHQ2.USTMA, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 19-JUL-2006, entry version 4.  
 DE 19-FEB-2006; entry version 1.  
 OS Hypothetical protein.  
 GN Unifam:UK00036.52;  
 OC Ustilaginaceae; Ustilaginomycetes;  
 OC Basidiomycota; Ustilaginomycetes;  
 OC Basidiomycota; Ustilaginomycetes; Ustilaginaceae; Ustilago.  
 NC [1]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=521;  
 RA Birren B.W., Nuebaum C., Abebe A., Abouelleil A., Adekoya E.,  
 RA Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,  
 RA Arachchi H.M., Armstrong P., Bachantang P., Baldwin J., Barry A.,  
 RA Bayul T., Blumsteyn B., Bloom T., Blye J., Boguslavskiy L.,  
 RA Borowsky M., Bouthalter B., Brunache A., Butler J., Calixte N.,  
 RA Calvo S.E., Camarata J., Campo K.C., Chang J., Cheshatsang Y.,  
 RA Citroen M., Collymore A., Considine T., Cook A., Cooke P., Corum B.,  
 RA Cuomo C., David R., Dawot T., Degray S., Dodge S., Dooley K.,  
 RA Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,  
 RA Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,  
 RA Fittsgerald M., Foley K., Gage D., Galagan J.E., Gearlson K., Gnerre S.,  
 RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafer N.,  
 RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,  
 RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,  
 RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysysselis M., Karlsson E.,  
 RA Kells C., Kieu A., Kisser P., Kodira C., Kulbokas E., Labutti K.,  
 RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,  
 RA Lindblad-Toh K., Liu X., Lokvisang T., Lokvisang Y., Lucien O.,  
 RA



Search completed: June 13, 2006, 14:04:46  
Job time : 321 secs



113 KYR-----NYNNGSI-----KDYKIKMDTNDSTNNHT-----HLAIPPIPIPTPII 156

109 QLPASAILTNSARPSPRLAHSMHPQOLYVESNI-----HTPPKDVGVKRDYTMSSSTA 161

157 TWANRSRKSOLENIPLPIKKYTIKGRNNINFDLVSPMKTKTNDSEDITNT----- 210

162 SSGNKS-KLSASSASPIKTVR-----KSSLVSPVLPIPHESKSDTHSKLAKPKK 210

211 ----STTAHKKLGICATTLGVGTGTATA-----TATAAGRRSPRSSITSEADSHAGRSQ 264

211 RTYSTTSAHSIN-PAVLLTKSTQSKDADDTLKERPKVRNTRAFSDSV-SQASRDSQ 268

265 ETEEDVCFPMVGD-HIRVNGIDFDEIFRIREREAYLQKQMLAK-NILRIDEFQNLSE 322

269 ETEEDVCFPMVQLTRVNGIDFDELEVAQANAE--KSQFLASLQVNEQKYSNVSQ 325

323 ----NNYTSAGSRHPYHHHNNKNGDGGSSMAALKYTPKNIKLLKTLGRFEFTH 376

326 DIGFTSTSTSGSS-----AALKYTPR--VSQGTGKSESTN 359

377 ENSSSSEBIYELKTKQOPPKYDDQLSTSTSTSGSGSGQVKGQ-ARISDINGSL 435

360 ET-----EIEHKKDEHEKIKPSLHPGFKNGKVEGENEIPSDNPAYCSVQGTDFOI 414

436 PDRSLPHSSEETIHAIDPISLVSPQSVDFLRNGEBETWMLDCTCTDSEMKMLAKAF 495

415 PNRFFCSDESETHASDTPSLVSEGOFTYELFRGCEPTWMLDSCPTDDDEMRCIAKAF 474

496 GIHPLTAEDIRMQETREKVELFKSYVFCFHTFEADKESDYLRPINVIYVFDHGLTF 555

475 GIHPLTAEDIRMQETREKVELFKSYVFCFHTFENDKESDPLPINVIYVCSGVLTF 534

556 HPSLSPHAINVRRVRLQDYVDYSADWLCYALIDETDGPAPVHIGIEYEAADAIDAYP 615

535 HFGPISHCANVRRVRLQDYVYVNSDWLCYALIDITDSFAPVIOSEIYEADAIEDSVF 594

616 TARDTDFSSMLQIGESRRKRWMLRLLSGKADVIKMFARKQBEANSSGSGYQORQNLQ 675

595 MAHMDMAFAMLRIGESRRKRWMLRLLSGKADVIKMFARKQBEANGL----- 643

676 QQQQQQPPPPPIITSPINSTLANLSGTSGSGVGVGINFGPNPQNNVNTNNTWTG 735

644 -----GALTSQIN-IANLQARQNA-----SHIKNNSSTVPNNY 678

736 SPSPQ0000GHITKSFPTDAPRADIALYGDIOHIIIMQNLLEYKIFSRSHN 795

679 APTSQ-----PRGIALYGDIOHIIIMQNLLEYKIFSRSHN 720

796 YLAQLOVESNSNKKITWFSKITGLWMLPLNLVGLPGNVRVVEGSGTNLGRFGFI 855

721 YLAQLOVESNSNKKITWGLVTHGLVPLAVITGLPGNVRVVEG--SSIAWMPFI 779

856 VGVLIPTIIGSIPAKWKKLNNSI-----EGQNGNRPFN-----HSSRSISLGLKK 907

780 LQVLLLVAVGWFLASWIKRIDPATLNEAESGAKSVISSFLPKRNRFPNDRSKNIV 839

908 HCG--NKSITSPNKKY 921

840 RAGPNSKVSAPLSRY 855

RESULT 2

S56205

probable membrane protein YFL050c - Yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein R001

C:Species: *Saccharomyces cerevisiae*

C>Date: 02-Sep-1995

C:Accession: S56205; S62287

R:Murekani, Y.; Naitou, M.; Hagiwara, M.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.;

Sasanuma, M.; Tsuchiya, Y.; Soeda, E.; Yokoyama, K.; Yamazaki, M.; Tashiro, H.; Eki,

T. submitted to the EMBL Data Library, May 1995

A:Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*.

A:Reference number: S56186

A:Accession: S56205

A:Molecule type: DNA

A:Residues: 1-858 <MUR>

615 TDAFVPLRIGIETETALESUVLGRSDSMLRRIGECRKTKTGWCPRLIIYKADVIKM 674  
 QY           |||||  
 653 FAKRCOEANSSGVYQRYNVLOQQOQAAPPPNPPIITSPINSLNLNSLGTSCTGCGVC 712  
 DB           |||||  
 675 LARCNENKIYA-----PFG-----EIGLYLDIQ 699  
 QY           |||||  
 713 VGGINFPGNPTGNNTWTNTWTCGSPSPPOOOQHGTNKSPFPIDPRADRIALVLGDIO 772  
 DB           |||||  
 687 -----FTG-----EIGLYLDIQ 699  
 QY           |||||  
 773 DLIITMFCNLLAYEKIFSRSHSNLYAOLVESFNPNKNKITEMFSKITLTCLTMVLVPNLVT 832  
 QY           |||||  
 700 DHLVTTWSLSQFKILSRTHSNLYAOLTSENISBRSRTGALKITLICTLLVAPNLVT 759  
 DB           |||||  
 833 GLFGNNVRVPEGGNTGLGWFFGVGLI-PIIGSFI 868  
 QY           |||||  
 760 GLFGNNVRVPEGRDNTNLAWFEGILGLVGSIAIGWII 796  
 DB           |||||

**RESULT 4**  
 T37833  
 probable membrane protein - fission yeast (*Schizosaccharomyces pombe*) (fragment)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
 C:Accession: T37833  
 R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
 submitted to the EMBL Data Library, September 1997  
 A:Reference number: Z21749  
 A:Accession: T37833  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-451 <MUR>  
 A:Cross-references: UNIPROT:O13779; UNIPARC:UPI0000169008; EMBL:Z29162;  
 PDB: CAB16212.1; GSPDB:GN00066; SPDB:SPAC17G6.01  
 A:Experimental source: strain 972H-; cosmid c17G6  
 C:Genetics:  
 A:Gene: SPAC17A2.14; SPDB:SPAC17G6.01  
 A:A:Map position: 1

Query Match 19.7%; Score 949; DB 2; Length 451;  
 Best Local Similarity 42.4%; Pred. No. 1.7e-50;  
 Matches 189; Conservative 64; Mismatches 87; Indels 106; Gaps 5;

QY 436 PDPSLFHSSEETHAPDISLVSPQSVDRLFNGEBTWMLDCCTPDISEMKKLAKAF 495 DB 101 PYRPFDFDSLPSTHSHEWEVLHDQSFEDLFHSG-GTNWLWDSCPKSEIRVLAKAF 159 QY 496 GHPLHTAEIDTMOQFSKVLKYSKYCPCHTFEADRESDEYLNYVLYPHDGILTLP 555 DB 160 GHPLTVEDITEERBEKVELRPYTPFRSPNQVPSNEVLPLNPLVPRDLITP 219 QY 556 HPESPISHANVRVRVOLRDYDVSDMWLCVALDELTCGPAPVHGIEYEADATE---- 611 DB 220 HNNPFPHPANVRERINQNLGYLVAUWKIALYALLDDYTDAFPTEIDEVDITDSMIL 279 QY 612 ----DAVFETARDTFSSMLORTGESRRKVMTWLLSKGADVHKFAKROCEANSSGV 667 DB 280 SIHYDEVNVEKPQE--RMQRVGECKRLINSLLRLANKADVVRGLSKRC----- 327 QY 668 YQRQYNLQQOQQQAAPPNPPIITSPINSLNLNSLGTSCTGCGVGCGVGFNPGNPTGN 727 DB 328 -----FTG-----EIGLYLDIQ 699 QY 728 NTWTWTCGSPSPPOOOQHGTNKSPFPIDPRADRIALVLGDIOQDIITMFONLLAYER 787 DB 328 -----NESMQV-----APRGETALYLDGVQDHIVTAVQNLNHYEK 362 QY 788 IFSRSHSNLYAOLVESFNPNKNKITEMFSKITLTCLTMVLVPNLVTGLFGNVRVPEGGT 847 DB 363 ILRSRSHSNLYAQISNNMTVLSNETNELSVLRSLTLATILIPNLVLTGLWGNVKVPGQDVP 422 QY 848 NLGWFFGVIGVLIPIIIGSFIAQWW 873 DB 423 GLCWFFSILGSLNIFAISSFLCKWY 448	615 TDAFVPLRIGIETETALESUVLGRSDSMLRRIGECRKTKTGWCPRLIIYKADVIKM 674 653 FAKRCOEANSSGVYQRYNVLOQQOQAAPPPNPPIITSPINSLNLNSLGTSCTGCGVC 712 675 LARCNENKIYA-----PFG-----EIGLYLDIQ 699 713 VGGINFPGNPTGNNTWTNTWTCGSPSPPOOOQHGTNKSPFPIDPRADRIALVLGDIO 772 687 -----FTG-----EIGLYLDIQ 699 773 DLIITMFCNLLAYEKIFSRSHSNLYAOLVESFNPNKNKITEMFSKITLTCLTMVLVPNLVT 832 700 DHLVTTWSLSQFKILSRTHSNLYAOLTSENISBRSRTGALKITLICTLLVAPNLVT 759 833 GLFGNNVRVPEGGNTGLGWFFGVGLI-PIIGSFI 868 760 GLFGNNVRVPEGRDNTNLAWFEGILGLVGSIAIGWII 796
--	---

**RESULT 5**  
 T37886  
 hypothetical protein YKL064w - yeast (*Saccharomyces cerevisiae*)

C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C:Accession: S37886  
R:Rasmussen, S.: von Wettstein, D.  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S37872  
A:Accession: S37886  
A:Molecule type: DNA  
A:Residues: 1-969 <RAN>  
A:Cross-references: UNIPROT:P35724; UNIPARC:UPI000012P2A4; ENBL:Z28064; NID:g486086;  
PIDD:CA481901.1; PID:g486087; MIPS:YKL064w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:MNR2  
A:Cross-references: SGD:S0001547; MIPS:YKL064w  
A:Map position: 11L

Query Match 17.9%; Score 865; DB 2; Length 969;  
Best Local Similarity 24.2%; Pred. No. 7e-45;  
Matches 271; Conservative 143; Mismatches 247; Indels 460; Gaps 29;

QY	11	STTNQPIPRSDVLDHNRNQITNDCAISDSEDELELKSELEVVKSE	-----	58
----	----	---	-------	----

Dbb 44 SDSRPTQLLDNLQHNHGQITDFDQI-DSWGMLHESDSTSDNDIIKSEDPSLKGAFIDHR 102

QY 59 ----KQQHHQEITSD-NAKPLTRKSGSSIKKSNLTDKDRITNPM5-LSGGDDTINSGH 112

Db 103 PSMSQPREGPQSVS TVQPQPI MKFSTPSYKKPAGLRPSDQNRSLVSDLSPSE--LESWL 160

QY 113 KNRNYMSSLRKDFYLKDN-TDDNSTNNHTHLAIPPIPTPIITNA-----NKSRRK 165

D<sub>b</sub> 161 KRR---KSVHKSFDENSEPTDRRQSNANDVIDVD-ALMNHVNNASTGVNDNSKRRK 215

QY 166 SQLENPLIKKKTIGRNNNSNFFENDLVSPMTKMKTNDSEDITNTSTTANHMKLGIGATT 225

Db 216 -----KKR--GSDSSN-KN--SKSTSSDSNDEEYNS----- 244

QY 226 LGVGTGTATATATAAGRRPSSRSDSEADSHASRSSQTEEDVCFPM--VGDIRVNG 283

Db 245 -----RPSS-----LSSNNSSLDVCLVLDDEGSEVPKAW 275

QY 284 IDPDEIDFIREEREEAYLQKQMIAKNILRIDEFQNL SKNNTTSGASRHPYHHHSNNKK 343

Db 276 PDCTVLEEPSKEETER--LRSQAI-----QDAEAFHFQYDEDEE 312

QY 344 NNGDGGSSMAALKYTPKNILKKTLSRFEFTHESSSEIYELKTKQPPKYDDQLS 403

Db 313 D-----GTSNEDGILFSPKPIVTNIDVPGLGNRRRVNETENLKNGLRKPRIAPWHLIQRP 367

QY 404 LTSSSTSGSGQVKGFGARISDGINGGSL-----PD--RFSLFHSES 446

Db 368 VL-----GSNSTKDSKSRIQSGLQDNLLVGRNIQYPPHIIISNNPEHFRFTYFRVDL 418

QY 447 EETIHAPDIPSLVSPGQSVRDLF----- 469

DB 419 DSTVHSPTISGLLQPGQRFQDLFVASIYSQDNSAGHIKTHPNSTPGIKAETVSQLQGLT 478

QY 470 -----RNGEET--WWLDCNCPDSEMMKLAFAIGHPLTAEDIRMQETREKVE 515

DB 479 ARNPFSTLSSMSVANIEDVPPFWLDVSNPTTEEMKILSKAFGIHPLTTEDIFLGEVREKVE 538

QY 516 LFRSYFVCFHTFE-----ADRES----- 534

DB 339 EFRWYILICFRSFDIVAERKHVRRRRKKEQESATLDHESISRRKKSQAYGATMSNESNANN 598

533 ----- 534

00 399 NSISNHSASAWLFSTLURAKRKSANKTININISSSSINRKRKVKSEKKKMEENEKFKRKSGRH 658

QY	--EDILEFINVIVFHDGLTIFHFSEISHFANVRKRVQRDRDVIDVSADWLCYALIDE	591
333	:    :    :   :    :   :    :   :	:

000 AFNEGEZEE ENVIATYKIOVUJIEHEAFIFHEINVRUKKERDIDENVISDWIAIALUD / 18

-----

-----

024

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible]

Db 332 TFIAGVYGNFKEMPE--LNSRWGYITWIVMLLIAGSLYFFMRKQWL 378

RESULT 7

H72360

divalent cation transport-related protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 31-Dec-2004

C:Accession: H72360

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.

A:Reference number: A72200; UID:99287316; PMID:10360571

A:Accession: H72360

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-351 <ARN>

A:Cross-references: UNIPROT:Q9WZ31; UNIPARC:UPI00000D39D8; GB:AE001731; GB:AE000512; NID:G4981074; PIDN:AAD35646.1; PID:G4981077; TIGR:TW0561

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0561

C:Superfamily: divalent cation transporter, Cora type

Query Match 5.28; Score 251; DB 2; Length 351;

Best Local Similarity 21.5%; Pred. No. 5.6e-08;

Matches 90; Conservative 63; Mismatches 131; Indels 134; Gaps 13;

Qy 469 FRNGSEWMLDCT-CPTDSEMKLAKAFGHPLTAEDIRMQTEKVELFKSYVFCFHT 527

Db 53 FRDSSTPTWINTGTHRTDVQVQGEFFGHPLVLEIDLNVHQRPKEFFENYVFLVKM 112

Qy 528 FPADESEEDVLPINVIYVFDHGLTFH-----FSPISHANVRVRQRDLRV-DVS 580

Db 113 FTYDKNLHE-LESEQVSLITKCNLFQEKIDVDFP-----VSRIRYNRGLIEKRK 165

Qy 581 ADMLCYALIDEITGCFAPVINGIEYEADEAFTARDTDFSSMLQIGESRRKVTLM 640

Db 166 ADYLYSLDALVDYVLEKIDDEIDVLESEVLERPEK-----TVQRTHLKKNLVLR 222

Qy 641 RLISGKADVTKFAKRCQBEANSSSGYQRYNQYQLQQQQQAPPPPPPIITSPINSTLN 700

Db 223 KTIWPLREVL-----SSLY-----RDVPL----- 242

Qy 701 NSLGSYTGSGGVGVGINFGNPNTGNNTWNTWNTGSPSPQOQOHHGNTKSFPIIDARP 760

Db 243 -----YFRDVTDTHTQIADVTETPRDIVSGLLDVYLSV-----SNKTNVNMKVLT 296

Qy 821 IGTMLVPLNLTGFGNVRVFGEGTNLQWFFG-----IVGLIFIIIGSFIPQW 872

Db 297 IATFMPLETFIAGIYGNFEPYMP-----LRWKGVPVLAIVGVAVVYVFKKKW 350

RESULT 8

H69806

divalent cation transport protein homolog yfjQ - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 31-Dec-2004

C:Accession: H69806

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grand, G.; Guiseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.;

Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E. A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takenaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambuto, R.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; UID:98044033; PMID:9384377

A:Accession: H69806

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-319 <KUN>

A:Cross-references: UNIPROT:Q31543; UNIPARC:UPI0000060088; GB:299108; GB:AL009126; NID:G2633055; PIDN:CAB12629.1; PID:G2633124

A:Experimental source: strain 168

C:Genetics:

A:Gene: yfjQ

C:Superfamily: divalent cation transporter, Cora type

Query Match 4.66; Score 223; DB 2; Length 319;

Best Local Similarity 17.7%; Pred. No. 2.5e-06;

Matches 80; Conservative 77; Mismatches 150; Indels 144; Gaps 12;

Qy 430 INGSGLDPRFSFHSESEETIAPDIPSLVSPQSQVRDLFRNGEETW-WLDCTCTPTDSEM 488

Db 2 INITAITTEHQLKKNIPIERVQPD-----AMWYVDFYGPDETT 42

Qy 489 KMLAKAFGHPLTAEDIRMQTEKVELFKSYVFCFHTPEADKESDELYEPINVIYV 548

Db 43 ALLRDFPHPLAIEDCFQMRPKLDHYGVYFVYHAL-----NKETLETERTVDFQ 97

Qy 549 HDGIUTFFHSPISHANVRVRQRDLRVQVDSADMLCALVIDEITDGFAPVINGIEYEA 608

Db 98 EKFFVYVFLHETPGIAKVRERLYASPOLKKGPHGISYIMDQLVDEYFPLVYKIEDRLN 157

Qy 609 ATEDAVTARDTDFSSMLQIGESRRKVTMLRLLSGKADVTKFAKRCQBEANSSSGY 668

Db 158 EIEE-----SRPHKTYGTQVLMNVEFDLTDLHLARTIIPWRL----- 196

Qy 669 QRQYNLQQQQQAPPPPPPIITSPINSTLNLSLGTSTGGGVGVGINFGNPNTGNNTN 728

Db 197 ---YRI-----LSLDRV----- 205

Qy 729 TWNTYTGSPSPQOQOHHGNTKSFPIIDARPADIALYLGIDIOHITTFQNLAYEKI 788

Db 206 -----KQRETKAYFSIDYDHLKLSEIVESNRDM 235

Qy 789 FGRSHSNYLAQLQVESFNKNITFMFSKITLIGTMLVPLNLTGFGNVRVFGEGTN 848

Db 236 TSDLADSYV-----TLNS-NRMAIWMVTUTVSTIPITPTFIAGVYGNFDFMPE----- 284

Qy 849 LGW---FFGIVGVLFIIIGSFIPAQWMLKK 876

Db 285 LHWKGYFVFLGMLAULVIGMLI---WFVHK 312

RESULT 9

S77053

magnesium/cobalt transport protein sli10671 - Synecocystis sp. (strain PCC 6803)

N:Alternate names: protein sli10671

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 31-Dec-2004

C:Accession: S77053

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A:Reference number: S74322; MUID:97061201; PMID:8905231  
 A:Accession: S77053  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-380 <KAN>  
 A:Cross-references: UNIPROT:Q55968; UNIPARC:UPI00000D344C; EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BAA10745.1; PID:g1006592  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C:Superfamily: divalent cation transporter, Cora type

Query Match 4.48; Score 214; DB 2; Length 380;  
 Best Local Similarity 19.28; Pred. No. 1.1e-05;  
 Matches 84; Conservative 72; Mismatches 141; Indels 140; Gaps 14;

```

Qy 478 LDCTCPTDSE-----MKLAKAFGIHPLTAEDIRMOETREKVELFKSYYP 522
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 59 LCCEAYVDTESVSWINIDGIGNHTWQGLGEVFKLHPVALEDVNVQREKVEYENHLI 118
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 523 VCFHTEADKESDYLFENIVVVFHDGILTFHESP--ISHPANVRVRQLRDYV-DVS 580
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 119 FTSRMVTLQSSQTFISEQISFILGKH-YLLTIQBEKPYDCLFSVRERIRTKGAIRQKN 177
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 581 ADWLCYALIDEITDGFAPVTHGIEYADAEVFTARDTDFSSMLQIGESRRKVTLM 640
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 178 ADYLFYALDAIIDGFFV---MEVYGLVQSLQSEIISCTPNKSLAKIHQLQDQILMR 234
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 641 RLISGKADVIKFAKRCQBEANSSGYQRYNLQOQQOQAPPPPPNPIITSPINSTLNL 700
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 235 RAIWPQRDAI----- 244
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 701 NSIGSTGGGVGGVGNFGNPTGNNTNTNTTSPSPQOQOQHGTNKSFPIDARP 760
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 245 NSL-----LRDGSDLISD--- 257
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 761 RADIALYIGDIQDHIITMFQNLAYEKIFSRSHSNLYLAQLOVESFNSNNKITEMFSKITL 820
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 258 --EYVFLRCDYDHTQIILDMETVRLASNLTDIYLSV-----SNMNEINKTLTV 308
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 821 IGMVLPLNIIVTGLFGNNVRVFGEGTN---LGMFFG--IVGVLFIIIGSFIAQW--- 872
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 309 IESIFIPITFIAGIYGNFN--PDKSPWNPMLNWTGYPVIVWVLTGVGMALYFVRKG 367
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 873 WKLKNNISREGQNNR 889
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 368 WFRNLND---VEKGNR 380
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10  
 E97128  
 magnesium and cobalt transport protein CAC1852 [imported] - Clostridium acetobutylicum  
 C:Species: Clostridium acetobutylicum  
 C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 31-Dec-2004  
 A:Accession: E97128  
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, H.M.; Dubois, J.; Qiu, D.; Hitti, J.; Wolf, Y.I.; Tatusov, R.L.; Sabathe, F.; Doucette-Stamm, L.; Soucaille, P.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*.  
 A:Reference number: A96900; MUID:21359325; PMID:21359325  
 C:Accession: E97128  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-354 <KUR>  
 A:Cross-references: UNIPROT:Q97104; UNIPARC:UPI00000CA31D; GB:AB001437; PIDN:AAK79816.1; PID:g15024829; GSPDB:GN00168  
 A:Experimental source: Clostridium acetobutylicum ATCC824  
 C:Genetics:  
 A:Gene: CAC1852  
 C:Superfamily: divalent cation transporter, Cora type

Query Match 4.48; Score 212.5; DB 2; Length 354;  
 Best Local Similarity 18.88; Pred. No. 1.3e-05;  
 Matches 89; Conservative 77; Mismatches 159; Indels 149; Gaps 15;

```

Qy 423 GARISDINGSLPDRFSLPHSESEET-IHAPDPSLVSPQSVRLD-----FRNGEETW 476
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 7 GYSKKGLKFGSL--IHVGNSTGKKTEIHIIDYKDDYFEKDVHSIDCKLYRDTDRMT 63
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 477 WLDCTCPTDSE--MKLAKAFGIHPLTAEDIRMOETREKVELFKSYFVCFHTEADKESK 535
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 64 WININGLEDIDLYEKVGKYGVHSLIMEDLNTNQRPKIEFEKVFIVLKMITYF--RE 120
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 536 DYLEPINVIYVFDHGLTFH-----FSPISHPANVRVRQLRDYV-DVSADMLCY 586
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 121 DKLWVEQVSIHCNKLITITFEEGKQGVF-----GNLDRIRMTKSKIRKKTGVGYLT 174
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 587 ALIDEITDGFAPVTHGIEYADAEVFTARDTDFSSMLQIGESRRKVTLMRLLSGK 646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 175 ALIDSIVDNTFVILEKVEDKIEFESLMT----- 204
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 647 ADVIKFAKRCQBEANSSGYQRYNLQOQQOQAPPPPPNPII---TSPINSTLNLS 702
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 205 -----TATNLPNFVYNLKRQ-----IFLWKAIVPLREIN--- 236
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 703 LGTSTGGGVGGVGNFGNPTGNNTNTNTTSPSPQOQOQHGTNKSFPIDARP 762
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 237 --TLQRGEVFIIG----- 248
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 763 DIALYIGDIQDHIITMFQNLAYEKIFSRSHSNLYLAQLOVESFNSNNKITEMFSKITL 822
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 249 DISIYFKDVTHTVQVITDTFELFRMTAGLLDTYLS-----SASNNKNEINKFLTIFS 301
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 823 TMLVPLNIIVTGLFGNNVRVFGEGTNLGMFFGIVGVLFIIIGSFIAQWMLK 876
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 302 TIFIPSLPVGVGYNFDMPE---LKFYGVAVLVWVIMISLSAFMLMTPFRKK 351
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11  
 AG2319  
 hypothetical protein all4110 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 31-Dec-2004  
 C:Accession: AG2319  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi, M.; Ishikawa, K.; Kimura, T.; Kishida, Y.; Kohara, M.; Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120.  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AG2319  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-380 <KUR>  
 A:Cross-references: UNIPROT:Q8YPTI; UNIPARC:UPI00000CEA09; GB:BA0000019; PIDN:BA875809.1; PID:g17133245; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all4110  
 C:Superfamily: divalent cation transporter, Cora type

Query Match 4.38; Score 207; DB 2; Length 380;  
 Best Local Similarity 20.08; Pred. No. 3e-05;  
 Matches 85; Conservative 71; Mismatches 126; Indels 144; Gaps 18;

```

Qy 477 WLDCTCPTDSE--MKLAKAFGIHPLTAEDIRMOETREKVELFKSYFVCFHTEADKES 534
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 74 WVDVQGLGNQDLERVGQVPELPLVLEDDVNVSRPKETDQDQLLFARWVVKERES 133
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 535 EYLEPINVIYVFDHGLTFHFSFISHP--ANVRVRQLRDYV-DVSADMLCYALIDE 591
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 134 GFYSEQVSS--LILGHKHYLLSQEEP-EHDCFAGVNRMRKGGKTRRQGGADYLAVALDA 190
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 592 ITDGFAPVTHGIEYADAEVFTARDTDFSSMLQIGESRRKVTLMRLLSGK 646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 191 IIDGFFV---LERYGALDLEEEVIVSPROT-----LQKIYHVRRELLQJRAIMPQ 242
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 647 ADVIKFAKRCQBEANSSGYQRYNLQOQQOQAPPPPPNPIITSPINSTLNLSLGS 706
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 243 RDAINAL-----IDRSQLISP----- 259
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```







```

QY 552 ILTFHSPISHPANRRVRQLRDYDVVSADWLCYALIDBITDGFAPVINGIEYEAIE 611
Db 541 -----LIDE-----GISENADSPF 554

QY 612 DAVPTARDYDFSSMLQRIGESRRKVTMLRLLSGKAD-----VIKMFAPR 656
Db 555 -----NLSEETQKINE---KIIELENLITSLNSNSNWSLNGSSTTISCNPLSPR 602

QY 657 COREANSSGYQRYNLQOQQQAPPPPNPI-ITSPINSTLNLSLGT---STGGGVG 712
Db 603 SKMPSSTSTSNITNSLRKFSQELKIELRPLDLRAELYSSINTSPRGSASISGSGSG 662

QY 713 VGGINFQPNPTGNNTNTNT-----NTGSPSPQOQQOQGITNKSPFIPDARPRAD 763
Db 663 GGGMNGCFKTSNSINSPIQFFENENESIDSYEKKNEQFESLTOLI-----RE 713

QY 764 IALYLGDIQDHIITMFQNLAYEKIFSRSHSNYLAQLQV-----ESFNSNNK 810
Db 714 NOLYTKPIE-----FREIKLLEKLESNSKSNIMQIEYKSTQLVLKQPKQDQSDKNIK 767

QY 811 ITEMFSKITLIGTMLVPLNLVTLGFGMNVRFEGGTLNGLWFFGIVGVLFIFIIGSFIPA 870
Db 768 RKQLFN-----GSNV-----SGSN-----781

QY 871 QMWLKKLNNSIEQONGNRPFPNHSRRSIRSLGLKKGONKSIISPPNKY 921
Db 782 -----NSGSGCGNHHHCNNS-----NGSNSEVI--PSKY 811

```

Search completed: June 13, 2006, 14:09:04  
Job time : 52 secs